

11-30-00

A

11/28/00
1c966US
11/28/00

IRON
N. Bernardo Avenue
Mountain View, California 94043
650.919.6637

Atty. Docket No. AVIR-0032-1

"Express Mail" Label No. EF 337287150

Date of Deposit 28 November 2000

PATENT APPLICATION
ASST. COMMISSIONER FOR PATENTS
Washington, D. C. 20231

I hereby certify that this is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Asst. Commissioner for Patents, Washington, D. C. 20231

Sir:

Transmitted herewith for filing is the

- ☒ patent application of
- ☐ design patent application of
- ☐ continuation-in-part patent application of

By Tracy J. D

Inventor(s): Kemble G.W., Duke, G.M., and Spaete, R.R.

For: ATTENUATION OF CYTOMEGALOVIRUS VIRULENCE

☐ This application claims priority from each of the following Application Nos./filing dates:
_____/_____; ____/____; ____/____.

☐ Please amend this application by adding the following before the first sentence: --This application claims the benefit of U.S. Provisional Application No. 60/____, filed ____.--

Enclosed are:

- ☒ 36 sheet(s) of ☐ formal ☐ informal drawing(s).
- ☐ An assignment of the invention to _____.
- ☒ A ☐ signed ☒ unsigned Declaration & Power of Attorney.
- ☐ A ☐ signed ☐ unsigned Declaration.
- ☐ A Power of Attorney.
- ☐ A verified statement to establish small entity status under 37 CFR 1.9 and 37 CFR 1.27 ☐ is enclosed ☐ was filed in the earliest of the above-identified patent application(s).
- ☐ A certified copy of a _____ application.
- ☐ Information Disclosure Statement under 37 CFR 1.97.
- ☐

In view of the Unsigned Declaration as filed with this application and pursuant to 37 CFR §1.53(d), Applicant requests deferral of the filing fee until submission of the Missing Parts of Application.

DO NOT CHARGE THE FILING FEE AT THIS TIME.

Telephone:
650.919.6637
APPNOFEE.TRN 4/96

Tracy J. D
Tracy J. Dunn
Reg. No.: 34,587
Attorney for Applicant

11/28/00
09/724935
U.S. P.O.

PATENT APPLICATION

ATTENUATION OF CYTOMEGALOVIRUS VIRULENCE

Inventor: George W. Kemble, a United States citizen
residing in Fremont, California

Gregory M. Duke, a United States citizen
residing in Palo Alto, California

Richard R. Spaete, a United States citizen
residing in Redwood City, California

Assignee: AVIRON

DOCKET "SEE 42460

ATTENUATION OF CYTOMEGALOVIRUS VIRULENCE

5

TECHNICAL FIELD

10 The present invention is related generally to methods and compositions for treating or preventing cytomegalovirus (CMV) infections, such as congenital CMV disease, CMV retinitis, CMV mononucleosis, and the like, and methods of attenuating pathogenic cytomegalovirus isolates and strains, genetically engineered cytomegaloviruses and combinations thereof, methods for altering the phenotype of CMV viruses, attenuated viral vaccine compositions, and uses thereof. More particularly, the present invention is related to methods and compositions for prophylaxis and therapy of human cytomegalovirus infection, including the use of methods that functionally inactivate a subset of cytomegalovirus genes present in pathogenic isolates of human cytomegalovirus.

BACKGROUND

20 Cytomegalovirus (CMV) is a widespread herpesvirus in the human population, with between 0.2 and 2.2% of the infant population becoming infected in utero and another 8-60% becoming infected during the first six months of life (Reynolds et al. (1973) New Engl. J. Med. 289: 1). Although CMV infections are most commonly subclinical, CMV-induced sensorineural hearing loss and fatal cytomegalovirus infections ("cytomegalic inclusion disease") are important public health problems. Moreover, CMV is one of the more common opportunistic infections associated with Acquired Immune Deficiency Syndrome ("AIDS") and frequently produces disease, with recurrent infection occurring in HIV-positive individuals, typically taking the form of retinitis or ulcerative lesions in the colon and esophagus, and occasionally producing extensive necrotization of the bowel with a grave prognosis (Rene et al. (1988) Dig. Dis. Sci. 33: 741; Meiselman et al. (1985) Gastroenterology: 88: 171). Cytomegalovirus (CMV) infection is the major infectious cause of mental retardation and congenital deafness. CMV is also responsible for a great deal

of disease among the immunosuppressed, producing general and often severe systemic effects in patients with AIDS, in organ transplant recipients who have been iatrogenically immunosuppressed, and in bone marrow transplant patients.

5 It is clear that cytomegalovirus infections are a significant human health problem. Therefore, it is desirable to develop prophylactic and therapeutic methods and compositions to prevent cytomegalovirus infection and/or inhibit recurrent infectious outbreaks from persistent latent infections,
10 particularly for treating CMV retinitis, CMV mononucleosis, and related CMV pathology in human patients.

One approach that has been used to treat herpesvirus infections is to inhibit CMV viral DNA replication. For example, viral DNA replication can frequently be inhibited by agents that inhibit virally-encoded DNA polymerase. The most notable examples of such inhibitors of viral DNA polymerase are acyclovir, ganciclovir, citrusine-I, and the acyclic guanosine phosphonate (R,S)-HPMPC (Terry et al. (1988) Antiviral Res. 10: 235; Yamamoto et al. (1989) Antiviral Res. 12: 21). However, these compounds
20 are not completely selective for viral thymidylate synthetases or DNA polymerases and therefore can disadvantageously cause inhibition of host DNA replication at high doses. Moreover, the development of mutant viruses which are resistant to the inhibitory effects of these compounds have been reported, and
25 appear to result from mutations in the viral DNA polymerase (Coen et al. (1982) J. Virol. 41: 909; Coen et al. (1980) Proc. Natl. Acad. Sci. (U.S.A.) 77: 2265; Larder et al. (1987) EMBO J. 6: 169). Thus, while CMV infections, such as CMV retinitis, can be initially treated with foscarnet and ganciclovir, after a period
30 of time CMV replication and progression of the pathological viral infection recurs.

Passive immunization with antibodies (e.g., immune globulin) has been tested in combination with ganciclovir for therapeutic efficacy in humans. Such antibody preparations are
35 obtained from the serum of donors, who possess a high antibody titre to the virus as a result of an earlier infection. One disadvantage of such conventional antibody preparations is the

limited number of suitable donors and the poor reproducibility or quality of the various preparations, including potential contamination with pathogens and pathogenic viruses. Unfortunately, the use of intravenous immune globulin in combination with ganciclovir apparently does not produce significantly improved efficacy as compared to ganciclovir treatment alone (Jacobson et al. (1990) Antimicrob. Agents and Chemother. 34: 176).

The safety and pharmacokinetic profiles of anti-cytomegalovirus monoclonal antibodies are discussed in Aulitzky et al. (1991) J. Infect. Dis. 163: 1344 and Drobyski et al. (1991) Transplantation 51: 1190. However, none of the reported human anti-CMV monoclonal antibodies have been shown to possess significant therapeutic efficacy in treating CMV infections (e.g., retinitis) in humans.

Attempts to use recombinantly produced hCMV glycoproteins as a subunit vaccine to provide protective immunity against hCMV infection and pathogenesis have not proven to be effective, but remain candidates for additional evaluation.

Thus, there exists a need in the art for effective methods and compositions for inhibiting human cytomegalovirus replication, attenuating CMV virulence in vivo, neutralizing CMV virions, and for preventing and treating human cytomegalovirus infections, and especially CMV infections in preborns, newborns, and immunosuppressed patients such as AIDS patients. For example but not limitation, a suitable attenuated human CMV vaccine which elicits satisfactory immunoprotection against CMV infection is needed in the art. The present invention fulfills these and other needs.

SUMMARY OF THE INVENTION

A basis of the present invention is the surprising and unexpected finding that: (1) clinical isolates of pathogenic CMV variants contain a genomic region ("virulence region") which typically is not present in CMV strains which have undergone extensive laboratory passaging of the virus in cell culture (hereafter termed "highly passaged strain variants") and (2)

functional disruption (e.g., deletion or insertional inactivation and the like) of genes in this genomic region produces a substantial attenuation of CMV virulence and/or pathogenicity in vivo. Furthermore, the virulence region of a clinical isolate of CMV is frequently deleted, rearranged, or substantially changed over the course of passaging the virus in cell culture.

In one aspect of the invention, the virulence region is obtained from an early passage Toledo strain and is conveniently termed the "Toledo genomic region" herein, although equivalent (e.g., homologous) regions or subsequences thereof are present in other clinical isolates of CMV besides the Toledo strain of CMV; the term "Toledo genomic region" encompasses these homologous regions in other clinical CMV isolates, many early passage CMV strains, and non-isolated pathogenic CMV variants.

The Toledo genomic region which is present in pathogenic CMV isolates and which is typically substantially absent in highly passaged CMV strains (e.g., AD169, high-passage Towne) has been sequenced and several open-reading frames have been identified (PCT Publication WO96/30387, U.S.S.N. 08/414,926, U.S.S.N. 08/644,543 filed 10 May 1996, each incorporated herein in their entirety by reference). Functional disruption of these open reading frames, either singly or in combination, has been unexpectedly found to substantially reduce virulence of the resultant CMV mutant(s) in vivo. Thus, in part, the invention provides methods and compositions for suppressing or inactivating expression of genes of the Toledo genomic region and its homolog regions in other CMV variants, and thereby reducing virulence and pathogenicity of clinically important CMV variants to generate a "Toledo region-attenuated CMV variant"; such Toledo region-attenuated CMV variants have altered phenotypes which generally make them candidates for use in live attenuated virus vaccines for prophylaxis and/or treatment of CMV disease. The invention is, in part, further based on the heretofore unrecognized finding that pathogenic clinical isolates of CMV have a distinct genome as compared to the commonly used laboratory-passaged strains of human CMV (e.g., AD169, highly-passaged Towne), and that the genomic region which is present in the clinical isolates and

which is substantially absent in laboratory-passaged strains confers enhanced virulence in vivo. Most common approaches to development of CMV therapies and vaccines have heretofore relied on laboratory-passaged strains which typically lack all or part of the Toledo genomic region and the genes encoded therein which have been unexpectedly found to confer enhanced in vivo virulence and are believed to contribute to clinical pathology and CMV-related disease.

The invention provides a method for attenuating virulence of CMV comprising functionally inactivating at least one open reading frame in a virulence region of a CMV genome having substantial identity to at least 300 bp, typically at least 500 bp, of a 15 kb sequence present in the genome of the Toledo strain of CMV and absent from the genome of the AD169 strain of CMV and/or absent from the genome of highly-passaged Towne (i.e., more than 50-100 passages). In an aspect, the method functionally inactivates at least one open reading frame present in a genomic region of a CMV genome having substantial identity to at least 300 bp of a 13 kb sequence present in the genome of the Toledo strain of CMV and absent from the genome of the Towne strain of CMV. In an embodiment, the method functionally inactivates at least one open reading frame present in a genomic region of a CMV genome having substantial identity to at least 500 bp of the sequence shown in Figs. 1A through 1T. In an embodiment, the method functionally inactivates at least the open reading frame corresponding to UL148 as identified herein. In a variation, the method functionally inactivates open reading frames in the region spanning UL138 to UL148. In an embodiment, the method functionally inactivates UL138, UL139, UL140, UL141, UL142, UL143, UL144, UL145, UL146, UL147, and/or UL148. In a variation, UL148 is inactivated singly or in combination with other open reading frames of the Toledo genomic region. In a specific embodiment, UL148 is inactivated in combination with UL141 and/or UL144. Typically, such Toledo region-attenuated CMV variants comprise at least 500 bp of the Toledo genomic region or a homolog region having at least 80 percent sequence identity; frequently they comprise at least 1.0

kbp of the Toledo genomic region or homolog virulence region; often they contain at least 5.0 kbp to 8.0 kbp of the Toledo genomic region or homolog virulence region, and can comprise up to a complete Toledo genomic region or homolog virulence region.

5 It is possible for a synthetic virulence region to be comprised of portions of two or more virulence regions (e.g., such as a chimeric virulence region comprising part of the Toledo genomic region from a first clinical isolate with a complementing portion of the Toledo genomic region of a second clinical isolate).

10 In an aspect, the invention provides a method for attenuating a CMV strain or isolate containing an encoding polynucleotide sequence encoding a polypeptide which is at least 80 percent sequence identical to a polypeptide encoded by UL138, UL139, UL140, UL141, UL142, UL143, UL144, UL145, UL146, UL147, and/or UL148 of the Toledo genomic region; the method comprising functionally inactivating (e.g., deleting or introducing a nonsense or missense mutation) said encoding polynucleotide sequence to produce a Toledo region-attenuated CMV variant. In a variation, all open reading frames (ORFs) in the CMV isolate that are at least 80% sequence identical to the corresponding sequence of the Toledo genomic region are functionally inactivated. In a variation, all open reading frames (ORFs) in the CMV isolate that are at least 80% sequence identical to UL138, UL139, UL140, UL141, UL142, UL143, UL144, UL145, UL146, UL147, and/or UL148 of the Toledo genomic region are functionally inactivated. In an alternate variation, only one or a subset of the open reading frames (ORFs) in the CMV isolate that are at least 80% sequence identical to the corresponding sequence(s) of the Toledo genomic region are functionally inactivated. Such Toledo region-attenuated CMV variants comprise at least 500 bp of a Toledo genomic region and can comprise up to a complete Toledo genomic region (including a chimeric Toledo genomic region composed from distinct clinical isolates or strains).

35 In an aspect, the invention provides a recombinant CMV virus, comprising a genome having at least 500 bp of a virulence region wherein at least one ORF has been functionally inactivated by a genetic alteration which is predetermined and/or which does

not occur in known isolates or strains of CMV regardless of passage history.

In an aspect, the method of attenuating virulence comprises functional inactivation of open reading frames by predetermined structural mutation (e.g., deletion, insertion, missense or nonsense mutation, and the like) of at least one open reading frame, or a predetermined mutation of a transcriptional control sequence that controls transcription of the open reading frame, or predetermined mutation of a splicing signal sequence or the like necessary for efficient expression of the encoded gene product of the open reading frame. In an embodiment, a selectable marker gene is introduced into an open reading frame, often in the portion of the open reading frame believed to encode the amino-terminal two-thirds of the gene product, to structurally disrupt the open reading frame and result in the inactivation of the open reading frame's capacity to encode its functional gene product. In a variation, open reading frame UL148 is structurally disrupted by mutation; in one embodiment the structural disruption results from insertion of a selectable and/or screenable marker gene (e.g., *gpt/lacZ*). In an embodiment, a selectable marker gene is used to replace all or part of at least one open reading frame, such as by replacement of a deleted region of the Toledo genomic region with a selectable marker gene. In a variation, a region spanning open reading frame UL138 to UL148 is structurally disrupted by mutation; in one embodiment the structural disruption results from deletion of the UL138-UL148 region and replacement with a selectable and/or screenable marker gene (e.g., *gpt/lacZ*).

In an aspect, the functional inactivation of a Toledo genomic region gene is provided by transcriptional and/or translational suppression with an antisense polynucleotide having a sequence of at least 15 nucleotides, typically at least 25 nucleotides, that are substantially complementary to a Toledo genomic region, most usually the antisense polynucleotide is substantially complementary to an open reading frame sequence of a Toledo genomic region open reading frame. In an embodiment, the antisense polynucleotide is substantially complementary to

at least 25 nucleotides of UL148. In an embodiment, the antisense polynucleotide is complementary to UL148 and further comprises additional 5' and/or 3' nucleotide(s) which are not substantially complementary to UL148. In variations, the antisense polynucleotides comprise non-natural chemical modifications, and can include, for instance, methylphosphonates, phosphorothioates, phosphoramidites, phosphorodithioates, phosphorotriesters, and boranophosphates. In a variation the antisense molecules can comprise non-phosphodiester polynucleotide analogs wherein the phosphodiester backbone is replaced by a structural mimic linkage include: alkanes, ethers, thioethers, amines, ketones, formacetals, thioformacetals, amides, carbamates, ureas, hydroxylamines, sulfamates, sulfamides, sulfones, and glycinydamides. In a variation, the invention provides peptide nucleic acids (PNAs) having a nucleobase sequence which is substantially complementary to a Toledo genomic region sequence, such as an open reading frame (e.g., UL148, UL141, UL142, etc.).

The invention also provides attenuated live virus CMV vaccines wherein at least one open reading frame of a Toledo genomic region is structurally disrupted. Typically, the UL148 open reading frame is structurally disrupted, either singly or in combination with other Toledo region open reading frames (e.g., UL141, UL144, and the like). Often the disruption of the open reading frame is an insertion, deletion, or replacement mutation which confers the property of reduced virulence as determined by a suitable in vivo virulence assay (e.g., see Experimental Examples). Toledo genomic region mutants which exhibit at least one log reduction, preferably two logs or more reduction, in virulence as determined by in vivo virulence assay, or other equivalent virulence measure, are attenuated CMV vaccines. Such attenuated CMV vaccines are used to immunize individuals to confer protective immunity, typically antibody-mediated and/or cell-mediated immunity, to prevent or reduce the severity of subsequent CMV infection following a suitable immunization period.

In an aspect, the invention also provides attenuated

live virus CMV vaccines wherein at least one open reading frame of a Toledo genomic region is replaced by a segment of the Towne genome which is not present in AD169. The Towne genome comprises a region not present in AD169; the region contains open reading frames designated UL147, UL152, UL 153, and UL154 and generally is spanned by nucleotides 178221 to 180029 of the Towne genome according to the AD169 numbering convention. An attenuated virus of the invention can, in one embodiment, comprise a Toledo genome wherein the Toledo genome region spanning open reading frames UL133 to UL151 are replaced with a Towne genome region spanning UL147, UL152, UL153, and UL154; this engineered CMV virus variant is an attenuated Toledo virus which comprises desirable features of Towne while reducing undesirable virulence of the Toledo genome region. The invention provides other variations of this basic method, whereby a segment of the Toledo genome region comprising at least one open reading frame is deleted or otherwise structurally disrupted in a CMV variant having a Toledo genome region or its homolog, and a segment of a Towne genome region comprising at least one open reading frame is inserted in the CMV variant. In an embodiment, the engineered CMV variant comprises: (1) Toledo DNA (DNA substantially identical to a Toledo strain, preferably identical to it) from about nucleotides 1 to about 168,000 corresponding to (i.e., according to) the AD169 nucleotide numbering convention, operably linked to (2) Towne DNA (DNA substantially identical to a Towne strain, preferably identical to it) from about nucleotides 143,824 to 189,466 according to the AD169 nucleotide numbering convention, operably linked to (3) Toledo DNA (DNA substantially identical to a Toledo strain, preferably identical to it) from about nucleotides 189,466 to about 209,514 corresponding to (i.e., according to) the AD169 nucleotide numbering convention, operably linked to (4) Towne DNA (DNA substantially identical to a Towne strain, preferably identical to it) from about nucleotides 200,080 to 229,354 according to the AD169 nucleotide numbering convention. The invention also provides vaccine compositions and formulations of such attenuated CMV viruses, which can include adjuvants, delivery vehicles, liposomal formulations, and the

like. The invention also provides the use of such attenuated CMV variants for prevention of CMV disease and infection; in one aspect this use includes administration of such vaccine to human subjects.

5 In a variation, the functional inactivation of a Toledo genomic region gene is provided by suppressing function of a gene product encoded by a Toledo region open reading frame by contacting or administering an antibody which is specifically reactive with said gene product. In an embodiment, the Toledo
10 genomic region gene is UL148, UL141, and/or UL144, typically at least UL148, although other Toledo open reading frames can be used. The antibody binds to a gene product encoded by a Toledo region open reading frame with an affinity of at least about $1 \times 10^7 \text{ M}^{-1}$, typically at least about $1 \times 10^8 \text{ M}^{-1}$, frequently at least
15 $1 \times 10^9 \text{ M}^{-1}$ to $1 \times 10^{10} \text{ M}^{-1}$ or more. In some aspects, the antibody is substantially monospecific. In an embodiment, the antibody is a human antibody raised by immunizing an individual with an immunogenic dose of a gene product of a Toledo region open reading frame. In an embodiment, the human antibody is a
20 monoclonal antibody, or collection of human monoclonal antibodies which bind to the Toledo region gene product(s). In an embodiment, the antibody is a humanized antibody comprising complementarity-determining regions substantially obtained from a non-human species immunoglobulin reactive with the Toledo
25 region gene product, and further comprising substantially human sequence framework and constant regions. The invention also comprises pharmaceutical formulations of such antibodies and the use of such antibodies to treat or prevent CMV diseases, such as by passive immunization or the like.

30 In an aspect, the invention provides a composite CMV variant comprising a highly-passaged Towne genome and at least one open reading frame of a Toledo genome region, typically present in or adjacent to the U_L/b' region of the composite CMV. In an aspect, the composite CMV is a highly-passaged Towne genome
35 further comprising a Toledo UL148, UL141, and/or UL144. In an embodiment, the composite CMV is a highly-passaged Towne genome with a complete Toledo genome region; in a variation said Toledo

genome region has at least one open reading frame functionally inactivated to further attenuate the virulence of the composite CMV. In a variation, a low passage Towne genome (i.e., less than 40 passages in culture) is used in place of a highly-passaged Towne genome. In an alternate variation, a virulence region from a low-passage Towne genome is emplaced in a Toledo genome so as to thereby replace at least 1 kpb of the virulence region of the Toledo genome with at least 500 bp, typically approximately the same length, of a corresponding region (e.g., substantial sequence identity) of low-passage Towne.

In an aspect, the invention provides a chimeric CMV virus, comprising a genome having a plurality of polynucleotide sequences, linked in conventional phosphodiester linkage, wherein at least two of said polynucleotide sequences are derived from different clinical isolates or strains of CMV. Said chimeric CMV virus can comprise a genome having a plurality of polynucleotide sequences, linked in conventional phosphodiester linkage, wherein a first CMV genome sequence of at least 500 bp and less than a complete CMV genome length (e.g., less than 250 kbp) is at least 98 percent sequence identical to a first CMV isolate or strain, and at least one additional CMV sequence of at least 500 bp and less than a complete CMV genome length (e.g., less than 250 kbp) is at least 98 percent sequence identical to a second CMV isolate or strain which has a genome having a polynucleotide sequence of at least 500 bp which is less than 60 percent sequence identical to any portion of the genome of said first CMV isolate or strain and/or which is absent or substantially absent in the genome of said first CMV isolate or strain. Said chimeric CMV virus comprises a genome having sufficient genetic information to replicate as a virus, typically as an infectious virus, in suitable host cells or a suitable host organism or replication system (e.g., SCID/hu thy/liv mice, human lung fibroblasts, and other systems known in the art). Generally, said chimeric CMV virus has a genome that comprises genetic information which is substantially sequence identical, generally at least 80 percent sequence identical, usually at least 95 percent sequence identical or more, to a high-passage Towne genome; said chimeric

CMV virus genome typically further comprises genetic information which is substantially sequence identical, generally at least 80 percent sequence identical, usually at least 95 percent sequence identical or more, to at least 1 kbp of a virulence region of a clinical isolate of CMV or a low-passage strain of CMV other than low-passage Towne; in an embodiment, a complete virulence region (e.g., Toledo genome region) of a clinical isolate or low-passage CMV strain is present.

In an aspect, the invention provides a chimeric CMV virus, comprising a chimeric genome comprising a polynucleotide having a first CMV sequence of at least 500 bp having at least 97 percent sequence identity with a genome of a first CMV isolate or CMV strain and a second CMV sequence of at least 500 bp having at least 97 percent sequence identity with a genome of a second CMV isolate or CMV strain, and wherein said chimeric genome comprises genetic information having substantial identity (e.g., at least 80 percent sequence identity, preferably at least 95 percent sequence identity) spanning at least about the complete low-passage Towne genome. Typically, the chimeric genome comprises at least 500 bp containing at least one ORF having at least 95 to preferably 100 percent sequence identity to a virulence region (e.g., Toledo genome region) of a clinical isolate or low-passage strain of CMV other than low-passage Towne.

In an aspect, the invention provides a chimeric CMV virus, comprising a chimeric genome comprising a polynucleotide having a first CMV sequence of at least 500 bp having at least 97 percent sequence identity with a genome of a first CMV isolate or CMV strain and a second CMV sequence of at least 500 bp having at least 97 percent sequence identity with a genome of a second CMV isolate or CMV strain, and wherein said chimeric genome comprises genetic information having substantial identity (e.g., at least 80 percent sequence identity, preferably at least 95 percent sequence identity) spanning at least about the complete Toledo genome excepting at least 1 kbp of the virulence determinign region of Toledo (Toledo genome region), and preferably excepting at least 5 kbp to the entire approximately

15kbp virulence-determining Toledo genome region. Typically, the chimeric genome comprises at least 500 bp containing at least one ORF having at least 95 to preferably 100 percent sequence identity to a virulence region of low-passage Towne.

5 In specific embodiments, the invention provides exemplary CMV chimeric viruses composed of genome portions of high-passage Towne and genome portions of Toledo; the exemplary CMV chimeric viruses are designated herein as Chimera I, Chimera II, Chimera III, Chimera IV, and Towne/Tol 11. In an aspect, the
10 invention encompasses these specific embodiments and variants of each exemplified Chimera wherein the boundaries (splice junctions/recombination joints) between the various Towne and Toledo genome portions vary from the specific exemplified Chimeras by less than 20 kbp, typically less than 10 kbp, usually
15 by less than 5 kbp, and in many embodiments by less than 1 kbp from the specific examples provided herein.

In a variation, the invention provides a diagnostic method for identifying a virulent CMV strain in a sample by detecting the presence of unique Toledo genome region
20 polynucleotide sequences and/or by detecting the presence of a polypeptide encoded by an open reading frame of the Toledo genomic region. Detection of polynucleotide sequences can be by any suitable method, including but not limited to PCR amplification using suitable primers, LCR, hybridization of a
25 labeled polynucleotide probe, and the like. Detection of polypeptide species is typically done by immunoassay using a specific antibody to the Toledo region gene product(s).

The invention also provides a method of treating or preventing CMV infection, the method comprising administering to
30 an individual an efficacious dose of a polypeptide which is substantially identical to the deduced amino acid sequence of UL148. In a variation, the polypeptide is a truncated variant, mutein, or analog of the deduced amino acid sequence of UL148, wherein the polypeptide is soluble.

35 A further understanding of the nature and advantages of the invention will become apparent by reference to the remaining portions of the specification and drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1R. Nucleotide sequence of Toledo genome region isolated from Toledo strain of HCMV.

Figures 2A-2H. Deduced amino acid sequences of open reading frames UL130 through UL151. Conventional single letter abbreviations are used.

Figure 3. Schematic representation of open reading frames and their location in Toledo genome region. Top line schematically portrays entire Toledo genome with U_L/b' region identified. Bottom line shows enlarged view of U_L/b' region. Arrows indicate polarity and length of open reading frame. Solid circles indicate potential glycosylation sites.

Figure 4. Schematic comparison of the novel genome regions of Toledo and highly-passaged Towne as compared to AD169.

Figure 5. CMV Towne and Toledo cosmids used to regenerate specific chimeric CMV viruses. The location of the cosmid insert are indicated beneath the appropriate viral genome. The numbers at the end of the insert denote the endpoints determined by DNA sequence analysis; the numbers correspond to AD169 genomic sequence in GenBank. "XXX" denotes an end which was refractory to DNA sequence analysis. These ends were mapped by restriction enzyme and Southern blot analyses. The vertical dashed line represents the location of the internal "a" sequence of the virus. The lower line depicts the structure of the Tol/Twn 39/50 genome. The thick gray line denotes sequences derived from Toledo and the thin black line depicts sequences contributed from highly-passaged Towne strain. Regions of overlap could be derived from either virus and are represented by a region of a thick gray and a thin black line together. The Tol/Twn 39/50 genome does not contain the Toledo genomic region.

Figure 6. Analysis of the gpt/LacZ recombinant viruses in the SCID-hu (thy/liv) model. Two independent isolates of Tol pGD6 and Tol pGD7 were tested in the model. 3 mice were used per group and the mean of the data is displayed. Error bars representing 2 standard errors from the mean are also displayed.

Figure 7. Southern blot showing that a variety of clinical isolates of CMV contain sequences homologous to the

Toledo U_L/b' region. The Towne lane contains genomic DNA from Aviron's highyl-passaged Towne strain (Towne AV).

Figure 8. Southern blot showing that previous variants of the Towne strain hybridize to the Toledo U_L/b' region.

5 Twn•Merck indicates Towne strain from the Merck clinical trial. Twn•MA, Twn•MA#5 and Twn•MA#8 are variants of Towne obtained from Microbiological Associates. Twn•Aviron is highly-passaged Towne obtained at Aviron.

10 Figure 9. Schematic depiction of generation of chimeric CMV virus genomes by cotransfection of cosmids containing portions of Towne and Toledo genomes.

15 Figure 10. Schematic depiction of the specific exemplary embodiments denoted Chimera I, Chimera II, Chimera III, Chimera IV, and Towne/Tol 11. Toledo genome is depicted as "Toledo"; highly-passaged Towne genome is depicted as "Towne•AV"; selected reading frames of importance, proposed function/homologues of selected ORFs, and scale (in kbp) is shown on the top line.

20 Figure 11. Replication of Toledo, highly passaged Towne, and Chimeras I, III, and IV (in order, respectively) in SCID-hu mice having a thymus/liver implant.

Figure 12. Schematic comparison of low-passage (long) Towne genome and high-passage (short) Towne genome.

25 Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or
30 equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. For purposes of the present invention, the following terms are defined below.

As used herein, the twenty conventional amino acids and
35 their abbreviations follow conventional usage (Immunology - A Synthesis, 2nd Edition, E.S. Golub and D.R. Gren, Eds., Sinauer Associates, Sunderland, Massachusetts (1991)). Stereoisomers

(e.g., D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α,α -disubstituted amino acids, N-alkyl amino acids, lactic acid, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ -carboxyglutamate, ϵ -N,N,N-trimethyllysine, ϵ -N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, ω -N-methylarginine, and other similar amino acids and imino acids (e.g., 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the righthand direction is the carboxy-terminal direction, in accordance with standard usage and convention. Similarly, unless specified otherwise, the lefthand end of single-stranded polynucleotide sequences is the 5' end; the lefthand direction of double-stranded polynucleotide sequences is referred to as the 5' direction. The direction of 5' to 3' addition of nascent RNA transcripts is referred to as the transcription direction; sequence regions on the DNA strand having the same sequence as the RNA and which are 5' to the 5' end of the RNA transcript are referred to as "upstream sequences"; sequence regions on the DNA strand having the same sequence as the RNA and which are 3' to the 3' end of the coding RNA transcript are referred to as "downstream sequences".

The term "naturally-occurring" as used herein as applied to an object refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring. Generally, the term naturally-occurring refers to an object as present in a non-pathological (undiseased) individual, such as would be typical for the species.

The term "corresponds to" is used herein to mean that a polynucleotide sequence is homologous (i.e., is identical, not strictly evolutionarily related) to all or a portion of a reference polynucleotide sequence, or that a polypeptide sequence

is identical to a reference polypeptide sequence. In contradistinction, the term "complementary to" is used herein to mean that the complementary sequence is homologous to all or a portion of a reference polynucleotide sequence. For illustration, the nucleotide sequence "TATAC" corresponds to a reference sequence "TATAC" and is complementary to a reference sequence "GTATA".

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity", and "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA or gene sequence given in a sequence listing, such as a polynucleotide sequence of Fig. 1A-1R, or may comprise a complete cDNA or gene sequence. A full-length cDNA or gene sequence is defined as a polynucleotide containing the sequence(s) necessary to encode a complete protein product, including a translation initiation codon and a translation termination codon, unless linked to another encoding sequence in a format for production as a fusion protein. Generally, a reference sequence is at least 20 nucleotides in length, frequently at least 25 nucleotides in length, and often at least 50 nucleotides in length. Since two polynucleotides may each (1) comprise a sequence (i.e., a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity.

A "comparison window", as used herein, refers to a conceptual segment of at least 20 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 20 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window

may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2: 482, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48: 443, by the search for similarity method of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. (U.S.A.) 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection, and the best alignment (i.e., resulting in the highest percentage of homology over the comparison window) generated by the various methods is selected.

The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The terms "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 25-50 nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of

the reference sequence over the window of comparison. The reference sequence may be a subset of a larger sequence, for example, as a segment of an open reading frame shown in Fig. 1A-1R.

5 As applied to polypeptides, the term "substantial identity" means that two peptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 80 percent sequence identity, preferably at least 90 percent sequence identity, more preferably at least
10 95 percent sequence identity or more (e.g., 99 percent sequence identity). Preferably, residue positions which are not identical differ by conservative amino acid substitutions.

Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For
15 example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having
20 aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are:
25 valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine.

The term "analog", "murein" or "mutant" as used herein refers to polypeptides which are comprised of a segment of at
30 least 10 amino acids that has substantial identity to a portion of the naturally occurring protein

The term "cognate" as used herein refers to a gene sequence that is evolutionarily and functionally related between species. For example but not limitation, in the human genome, the human CD4 gene is the cognate gene to the mouse CD4 gene,
35 since the sequences and structures of these two genes indicate that they are highly homologous and both genes encode a protein which functions in signaling T cell activation through MHC class

II-restricted antigen recognition.

The term "agent" is used herein to denote a chemical compound, a mixture of chemical compounds, an array of spatially localized compounds (e.g., a VLSIPS peptide array, polynucleotide array, and/or combinatorial small molecule array), a biological macromolecule, a bacteriophage peptide display library, a bacteriophage antibody (e.g., scFv) display library, a polysome peptide display library, or an extract made from biological materials such as bacteria, plants, fungi, or animal (particularly mammalian) cells or tissues. Agents are evaluated for potential activity as antineoplastics, anti-inflammatories, or apoptosis modulators by inclusion in screening assays described hereinbelow. Agents are evaluated for potential activity as specific protein interaction inhibitors (i.e., an agent which selectively inhibits a binding interaction between two predetermined polypeptides but which does not substantially interfere with cell viability) by inclusion in screening assays.

As used herein, the terms "label" or "labeled" refers to incorporation of a detectable marker, e.g., by incorporation of a radiolabeled amino acid or attachment to a polypeptide of biotinyl moieties that can be detected by marked avidin (e.g., streptavidin containing a fluorescent marker or enzymatic activity that can be detected by optical or calorimetric methods). Various methods of labeling polypeptides and glycoproteins are known in the art and may be used. Examples of labels for polypeptides include, but are not limited to, the following: radioisotopes (e.g., ^3H , ^{14}C , ^{35}S , ^{125}I , ^{131}I), fluorescent labels (e.g., FITC, rhodamine, lanthanide phosphors), enzymatic labels (e.g., horseradish peroxidase, β -galactosidase, luciferase, alkaline phosphatase), biotinyl groups, predetermined polypeptide epitopes recognized by a secondary reporter (e.g., leucine zipper pair sequences, binding sites for secondary antibodies, transcriptional activator polypeptide, metal binding domains, epitope tags). In some embodiments, labels are attached by spacer arms of various lengths to reduce potential steric hindrance.

As used herein, "substantially pure" means an object

species is the predominant species present (i.e., on a molar basis it is more abundant than any other individual macromolecular species in the composition), and preferably a substantially purified fraction is a composition wherein the object species comprises at least about 50 percent (on a molar basis) of all macromolecular species present. Generally, a substantially pure composition will comprise more than about 80 to 90 percent of all macromolecular species present in the composition. Most preferably, the object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered macromolecular species.

The term "primer" as used herein refers to an oligonucleotide whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, i.e., in the presence of nucleotides and an agent for polymerization such as DNA polymerase and at a suitable temperature and pH. The primer is preferably single-stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the agent for polymerization. The exact lengths of the primers will depend on many factors, including temperature and source of primers. For example, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15-25 or more nucleotides, although it may contain fewer nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with template. In some embodiments, the primers can be large polynucleotides, such as

from about 200 nucleotides to several kilobases or more. The primers herein are selected to be substantially complementary to the different strands of each specific sequence to be amplified. The primers must be sufficiently complementary to hybridize with
5 their respective strands. Therefore, the primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment may be attached to the 5' end of the primer, with the remainder of the primer sequence being complementary to the strand. Alternatively,
10 noncomplementary bases or longer sequences can be interspersed into the primer, provided that the primer sequence has sufficient complementarity with the sequence of the strand to be amplified to hybridize therewith and thereby form a template for synthesis of the extension product of the other primer.

00321" 5642450
15 The term "recombinant" used herein refers to macromolecules produced by recombinant DNA techniques wherein the gene coding for a polypeptide is cloned by known recombinant DNA technology. For example, an amplified or assembled product polynucleotide may be inserted into a suitable DNA vector, such
20 as a bacterial plasmid, and the plasmid used to transform a suitable host. The gene is then expressed in the host to produce the recombinant protein. The transformed host may be prokaryotic or eukaryotic, including mammalian, yeast, Aspergillus and insect cells. One preferred embodiment employs bacterial cells as the
25 host. Alternatively, the product polynucleotide may serve a non-coding function (e.g., promoter, origin of replication, ribosome-binding site, etc.).

30 DETAILED DESCRIPTION

Commonly-assigned U.S. patent application U.S.S.N. 08/414,926 filed 31 March 1995 is incorporated herein by reference.

35 The nomenclature used hereafter and the laboratory procedures in cell culture, molecular genetics, and nucleic acid chemistry and hybridization described below may involve well known and commonly employed procedures in the art. Standard

techniques are used for recombinant nucleic acid methods, polynucleotide synthesis, and microbial culture and transformation (e.g., electroporation, lipofection). The techniques and procedures are generally performed according to conventional methods in the art and various general references (see, generally, Sambrook et al. Molecular Cloning: A Laboratory Manual, 2d ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Oligonucleotides can be synthesized on an Applied Bio Systems oligonucleotide synthesizer according to specifications provided by the manufacturer.

Methods for PCR amplification are described in the art (PCR Technology: Principles and Applications for DNA Amplification ed. HA Erlich, Stockton Press, New York, NY (1989); PCR Protocols: A Guide to Methods and Applications, eds. Innis, Gelfland, Snisky, and White, Academic Press, San Diego, CA (1990); Mattila et al. (1991) Nucleic Acids Res. 19: 4967; Eckert, K.A. and Kunkel, T.A. (1991) PCR Methods and Applications 1: 17; and U.S. Patent Nos. 4,683,202 and 4,965,188, each of which are incorporated herein by reference) and exemplified hereinbelow.

It is evident that optimal PCR and hybridization conditions will vary depending upon the sequence composition and length(s) of the targeting polynucleotide(s) and target(s), and the experimental method selected by the practitioner. Various guidelines may be used to select appropriate primer sequences and hybridization conditions (see, Maniatis et al., Molecular Cloning: A Laboratory Manual (1989), 2nd Ed., Cold Spring Harbor, N.Y.; Berger and Kimmel, Methods in Enzymology, Volume 152, Guide to Molecular Cloning Techniques (1987), Academic Press, Inc., San Diego, CA; PCR Protocols: A Guide to Methods and Applications, eds. Innis, Gelfland, Snisky, and White, Academic Press, San Diego, CA (1990); Benton WD and Davis RW (1977) Science 196: 180; Goodspeed et al. (1989) Gene 76: 1; Dunn et al. (1989) J. Biol. Chem. 264: 13057 which are incorporated herein by reference.

A basis of the invention is the unexpected discovery

that there are significant genomic differences between clinical isolates of CMV and highly-passaged CMV strains, including differences between low-passage Towne and high-passage Towne, as well as differences as compared to Toledo strain; the identification of these genomic differences, including definition of novel genomic region(s); and the phenotypic significance and biological function of said genomic differences and specific ORFs within said novel genomic regions. Based, in part, on these unexpected discoveries, it is possible to construct and use chimeric CMV viruses which have predetermined genome compositions comprising at least a portion of a genome of a first CMV isolate or strain and at least a portion of a genome of a second (or subsequent) CMV isolate or strain, so as to form a complete, replicable recombinant chimeric CMV genome, with and the resultant chimeric CMV genome being capable of replication in a suitable host replication system and being useful for a variety of uses, such as human or veterinary vaccines, commercial reagents for laboratory use (e.g., as restriction enzymes are sold), use in screening systems to identify novel candidate drugs to inhibit replication or pathogenesis (e.g., virulence, tropism, host range, etc.) of pathogenic, clinically relevant CMV virus types, and other uses such as diagnostic reagents, gene expression vectors, anti-tumor agents, heterologous gene expression systems, and the like.

Overview

An approach of the invention starts with identification of DNA sequences which confer virulence on human cytomegalovirus (HCMV). These sequences can be manipulated to produce a new, more efficacious HCMV vaccine strain with predicted characteristics. Introduction of the virulence genes into an overattenuated strain can improve its immunogenicity and deletion of the virulence genes from a virulent strain can render it safe in humans by decreasing its virulence. Specifically, deletion of genetic information from a clinical isolate called Toledo is used to attenuate an HCMV virus, and in one embodiment, a segment from a laboratory strain called Towne, especially a highly-

passaged Towne variant, is transferred to the deleted region of Toledo to act as a "spacer". Deleting genetic information has utility in improving a clinical isolate such as Toledo as an immunizing composition. Removing these sequences from Toledo, which has been shown to cause disease in people, can result in an attenuated virus which may be a safe vaccine candidate.

10 The Towne strain of HCMV has been used as a vaccine in humans. In some clinical settings, Towne has been used to prevent the disease consequences associated with infection by HCMV (reviewed by Marshall and Plotkin In: *The Human Herpesviruses* B. Roizman, R.J. Whitley, & C. Lopez Eds. Raven Press, New York). The Towne strain is believed to be overattenuated as a vaccine candidate and consequently, is poorly immunogenic. This loss of immunogenicity may have been the result of an extensive passage history in tissue culture. Genetic information in the virulence region may have been lost during passage, particularly after about Passage 40. Variation in DNA content among isolated strains does exist based on crude hybridization experiments. Other investigators have reported minor regions of sequence heterogeneity between two so-called laboratory strains of HCMV, the Towne and AD169 strains. Heterogeneities can exist within HCMV strains depending upon the extent of passages in their culture history.

25 The public health impact of HCMV infections have not been well controlled by current treatment strategies or available antiviral chemotherapies. Preventive vaccine strategies are likely to prove efficacious because of the observations that seropositive renal allograft recipients are protected from severe HCMV disease and maternal immunity protects the fetus from disease after intrauterine infection. HCMV (Towne) was developed as a vaccine strain by serial passage 125 times in WI38 human diploid fibroblasts (Towne 125). It has been administered to humans without significant adverse reactions. However, in one study, vaccinees were directly challenged by wild-type virus and found to resist only low challenge doses of 10 plaque-forming units or less. The consensus view is that the Towne strain may be overly attenuated. One positive feature of the Towne strain

is that it has never been shown to reactivate.

One important obstacle to the development of a vaccine for HCMV is the lack of an animal model system that can be used to test the safety and efficacy of vaccine candidates. Therefore, cell culture systems or surrogate animal models such as the SCID-hu (thy/liv) mouse have to be developed to test vaccine strains. Replicative differences in HCMV strains have been described in a variety of cell types and in the SCID-hu mouse model. These differences correlate to the virulence and passage history of the virus. Thus, low passage, virulent clinical isolates, such as Toledo, can replicate better in the human implant of SCID-hu (thy/liv) mice and in cultures of human endothelial cells than cell culture adapted, highly-passaged avirulent laboratory strains such as Towne or AD169 (Brown et al. 1995; Waldman et al., 1991). This observation can be exploited to measure the "virulence" of a strain by assessing its growth characteristics in the SCID-hu mouse, in vivo in humans, or by other means. Recombinant vaccine candidates such as the ones described here which have deleted or incorporated DNA sequences are believed to replicate less well than the virulent parent in a suitable virulence assay. This observation would be indicative of an attenuated vaccine candidate. Deletion of the Toledo UL/b' region from the low passage, virulent HCMV Toledo genome results in a virus with reduced replicative ability in the SCID-hu mouse. This recombinant virus should have a concomitantly reduced virulence which allows administration of the virus without causing the undesired clinical manifestations exhibited by the Toledo virus in humans.

The invention identifies, maps, and sequences differences between the virulent Toledo strain and the avirulent highly passaged Towne strain, for the purpose of transferring novel genetic information to Towne to restore its immunogenicity or, alternatively, to remove information from Toledo to render it safe as a vaccine candidate. One major region of difference mapped to the internal portion of the L component. This large 13kbp region present in Toledo but not highly passaged Towne is located at the border between the unique long (UL) and the

inverted repeats bordering the UL region termed IRL or b'. We have deduced the coding information resident in the Toledo sequences and have extensively compared the information resident in AD169, highly passaged Towne and Toledo. We have made recombinant viruses which have either inserted the UL/b' region from the virulent Toledo strain, into the corresponding region of Towne, and have also deleted this region from Toledo and replaced it with a selectable marker and reporter gene or with the corresponding UL/b' region from Towne. Deletion of the virulence genes from Toledo decreased the ability of the recombinant to replicate within the SCID-hu (thy/liv) mouse, a model for CMV virulence. The new recombinant viruses exhibit growth properties in the SCID-hu mouse that indicate that vaccine candidates with attenuated virulence can be generated by deleting the UL/b' region from the Toledo virus. We have also demonstrated that we can add the Toledo region to the Towne virus which will presumably result in increased immunogenicity for the highly passaged Towne virus while retaining its safe profile for humans.

Figures 1A-1R show the nucleotide sequence of Toledo genome region isolated from Toledo strain of HCMV. Figures 2A-2H show the deduced amino acid sequences of open reading frames UL130 through UL151.

A basis of the present invention is the surprising and unexpected finding that: (1) clinical isolates of pathogenic CMV variants contain a genomic region which typically is not present in CMV strains which have undergone extensive laboratory passaging of the virus in cell culture, and (2) functional disruption (e.g., deletion or insertional inactivation and the like) of genes in this genomic region produces a substantial attenuation of CMV virulence and pathogenicity in vivo. The genomic region is conveniently termed the "Toledo genomic region" herein, although equivalent (e.g., homologous) regions or subsequences thereof are present in other clinical isolates of CMV besides the Toledo strain of CMV; the term "Toledo genomic region" encompasses these homologous regions in other clinical CMV isolates and non-isolated pathogenic CMV variants which have a genomic region of at least 500 bp having at least 80 percent

sequence identity to the Toledo genomic region of the Toledo strain having the sequences disclosed herein and in WO96/30387, incorporated herein by reference. The Toledo genomic region which is present in pathogenic CMV isolates and which is typically substantially absent in laboratory passaged CMV strains (e.g., AD169, Towne) has been sequenced and several open-reading frames have been identified. Functional disruption of these open reading frames, either singly or in combination, has been unexpectedly found to substantially reduce virulence of the resultant CMV mutant(s) in vivo. Thus, in part, the invention provides methods and compositions for suppressing or inactivating expression of genes of the Toledo genomic region and its homolog regions in other CMV variants, and thereby reducing virulence and pathogenicity of clinically important CMV variants. The invention is, in part, further based on the heretofore unrecognized finding that pathogenic clinical isolates of CMV have a distinct genome as compared to the commonly used laboratory-passaged strains of human CMV (e.g., AD169, Towne), and that the genomic region which is present in the clinical isolates and which is substantially absent in laboratory-passaged strains confers enhanced virulence in vivo. Most common approaches to development of CMV therapies and vaccines have heretofore relied on laboratory-passaged strains which lack the Toledo genomic region and the genes encoded therein which have been unexpectedly found to confer enhanced in vivo virulence and are believed to contribute to clinical pathology and CMV-related disease.

The invention provides a method for attenuating virulence of CMV comprising functionally inactivating at least one open reading frame in a genomic region of a CMV genome having substantial sequence identity to at least 300 bp, typically at least 500 bp, of an approximately 15 kb sequence present in the genome of the Toledo strain of CMV and absent from the genome of the AD169 strain of CMV. In an aspect, the method functionally inactivates at least one open reading frame present in a genomic region of a CMV genome having substantial identity to at least 300 bp of a 13 kb sequence present in the genome of the Toledo

strain of CMV and absent from the genome of the highly -passaged Towne strain of CMV. In an embodiment, the method functionally inactivates at least one open reading frame present in a genomic region of a CMV genome having substantial identity to at least 500 bp of the sequence shown in Figs. 1A through 1T. In an embodiment, the method functionally inactivates at least the open reading frame corresponding to UL148 as identified herein. In a variation, the method functionally inactivates open reading frames in the region spanning UL138 to UL148. In an embodiment, the method functionally inactivates UL138, UL139, UL140, UL141, UL142, UL143, UL144, UL145, UL146, UL147, and/or UL148. In a variation, UL148 is inactivated singly or in combination with other open reading frames of the Toledo genomic region. In a specific embodiment, UL148 is inactivated in combination with UL141 and/or UL144. Inactivation is typically accomplished by genetic engineering and involves predetermined mutations (which may include additions, transpositions, or deletions), generally of the specific type which are not known to occur naturally in CMV strains even after extensive passaging.

In an aspect, the method of attenuating virulence comprises functional inactivation of open reading frames by structural mutation (e.g., deletion, insertion, missense or nonsense mutation, and the like) of at least one open reading frame, or a mutation of a transcriptional control sequence that controls transcription of the open reading frame, or mutation of a splicing signal sequence or the like necessary for efficient expression of the encoded gene product of the open reading frame. In an embodiment, a selectable marker gene is introduced into an open reading frame, often in the portion of the open reading frame believed to encode the amino-terminal two-thirds of the gene product, to structurally disrupt the open reading frame and result in the inactivation of the open reading frame's capacity to encode its functional gene product. In a variation, open reading frame UL148 is structurally disrupted by predetermined mutation, often produced by site-directed mutagenesis or in vitro recombination; in one embodiment the structural disruption results from insertion of a selectable and/or screenable marker

gene (e.g., *gpt/lacZ*). In an embodiment, a selectable marker gene is used to replace all or part of at least one open reading frame, such as by replacement of a deleted region of the Toledo genomic region with a selectable marker gene. In a variation, a region spanning open reading frame UL138 to UL148 is structurally disrupted by predetermined mutation; in one embodiment the structural disruption results from deletion of the UL138-UL148 region and replacement with a selectable and/or screenable marker gene (e.g., *gpt/lacZ*).

In an aspect, the functional inactivation of a Toledo genomic region gene is provided by transcriptional and/or translational suppression with an antisense polynucleotide having a sequence of at least 15 nucleotides, typically at least 25 nucleotides, that are substantially complementary to a Toledo genomic region, most usually the antisense polynucleotide is substantially complementary to an open reading frame sequence of a Toledo genomic region open reading frame. In an embodiment, the antisense polynucleotide is substantially complementary to at least 25 nucleotides of UL148. In an embodiment, the antisense polynucleotide is complementary to UL148 and further comprises additional 5' and/or 3' nucleotide(s) which are not substantially complementary to UL148. In variations, the antisense polynucleotides comprise non-natural chemical modifications, and can include, for instance, methylphosphonates, phosphorothioates, phosphoramidites, phosphorodithioates, phosphorotriesters, and boranophosphates. In a variation the antisense molecules can comprise non-phosphodiester polynucleotide analogs wherein the phosphodiester backbone is replaced by a structural mimic linkage include: alkanes, ethers, thioethers, amines, ketones, formacetals, thioformacetals, amides, carbamates, ureas, hydroxylamines, sulfamates, sulfamides, sulfones, and glycinydamides. In a variation, the invention provides peptide nucleic acids (PNAs) having a nucleobase sequence which is substantially complementary to a Toledo genomic region sequence, such as an open reading frame (e.g., UL148, UL141, UL142, etc.).

The invention also provides attenuated live virus CMV

vaccines wherein at least one open reading frame of a Toledo genomic region is structurally disrupted by predetermined mutation. Typically, the UL148 open reading frame is structurally disrupted, either singly or in combination with other Toledo region open reading frames (e.g., UL141, UL144, and the like). Often the disruption of the open reading frame is an insertion, deletion, or replacement mutation which confers the property of reduced virulence as determined by a suitable in vivo virulence assay (e.g., see Experimental Examples). Toledo genomic region mutants which exhibit at least one log reduction, preferably two logs or more reduction, in virulence as determined by in vivo virulence assay, or other equivalent virulence measure, are attenuated CMV vaccines. Such attenuated CMV vaccines are used to immunize individuals to confer protective immunity, typically antibody-mediated and/or cell-mediated immunity, to prevent or reduce the severity of subsequent CMV infection following a suitable immunization period.

In an aspect, the invention also provides attenuated live virus CMV vaccines wherein at least one open reading frame of a Toledo genomic region is replaced by a segment of the Towne genome which is not present in AD169. The highly-passaged Towne genome comprises a region not present in AD169; the region contains open reading frames designated UL147, UL152, UL 153, and UL154 and generally is spanned by nucleotides 178221 to 180029 of the Towne genome according to the AD169 numbering convention. An attenuated virus of the invention can, in one embodiment, comprise a Toledo genome wherein the Toledo genome region spanning open reading frames UL133 to UL151 are replaced with a Towne genome region spanning UL147, UL152, UL153, and UL154; this engineered CMV virus variant is an attenuated Toledo virus which comprises desirable features of Towne while reducing undesirable virulence of the Toledo genome region. The invention provides other variations of this basic method, whereby a segment of the Toledo genome region comprising at least one open reading frame is deleted or otherwise structurally disrupted in a CMV variant having a Toledo genome region or its homolog, and a segment of a Towne genome region comprising at least one open reading frame

Smith Barney
vestrierte stock
212-723 9750
Wayne Brontstein

in inserted in the CMV variant. In an embodiment, the engineered CMV variant comprises: (1) Toledo DNA (DNA substantially identical to a Toledo strain, preferably identical to it) from about nucleotides 1 to about 168,000 corresponding to (i.e., according to) the AD169 nucleotide numbering convention, operably linked to (2) Towne DNA (DNA substantially identical to a Towne strain, preferably identical to it) from about nucleotides 143,824 to 189,466 according to the AD169 nucleotide numbering convention, operably linked to (3) Toledo DNA (DNA substantially identical to a Toledo strain, preferably identical to it) from about nucleotides 189,466 to about 209,514 corresponding to (i.e., according to) the AD169 nucleotide numbering convention, operably linked to (4) Towne DNA (DNA substantially identical to a Towne strain, preferably identical to it) from about nucleotides 200,080 to 229,354 according to the AD169 nucleotide numbering convention. The invention also provides vaccine compositions and formulations of such attenuated CMV viruses, which can include adjuvants, delivery vehicles, liposomal formulations, and the like. The invention also provides the use of such attentuated CMV variants for prevention of CMV disease and infection; in one aspect this use includes administration of such vaccine to human subjects.

In a variation, the functional inactivation of a Toledo genomic region gene is provided by suppressing function of a gene product encoded by a Toledo region open reading frame by contacting or administering an antibody which is specifically reactive with said gene product. In an embodiment, the Toledo genomic region gene is UL148, UL141, and/or UL144, typically at least UL148, although other Toledo open reading frames can be used. The antibody binds to a gene product encoded by a Toledo region open reading frame with an affinity of at least about $1 \times 10^7 \text{ M}^{-1}$, typically at least about $1 \times 10^8 \text{ M}^{-1}$, frequently at least $1 \times 10^9 \text{ M}^{-1}$ to $1 \times 10^{10} \text{ M}^{-1}$ or more. In some aspects, the antibody is substantially monospecific. In an embodiment, the antibody is a human antibody raised by immunizing an individual with an immunogenic dose of a gene product of a Toledo region open reading frame. In an embodiment, the human antibody is a

monoclonal antibody, or collection of human monoclonal antibodies which bind to the Toledo region gene product(s). In an embodiment, the antibody is a humanized antibody comprising complementarity-determining regions substantially obtained from
5 a non-human species immunoglobulin reactive with the Toledo region gene product, and further comprising substantially human sequence framework and constant regions. The invention also comprises pharmaceutical formulations of such antibodies and the use of such antibodies to treat or prevent CMV diseases, such as
10 by passive immunization or the like.

In an aspect, the invention provides a composite CMV variant comprising a Towne genome and at least one open reading frame of a Toledo genome region, typically present in or adjacent to the U_L/b' region of the composite CMV. In an aspect, the
15 composite CMV is a Towne genome further comprising a Toledo UL148, UL141, and/or UL144. In an embodiment, the composite CMV is a highly-passaged Towne genome with a complete Toledo genome region; in a variation said Toledo genome region has at least one open reading frame functionally inactivated to further attenuate the virulence of the composite CMV.
20

In a variation, the invention provides a diagnostic method for identifying a virulent CMV strain in a sample by detecting the presence of unique Toledo genome region polynucleotide sequences and/or by detecting the presence of a
25 polypeptide encoded by an open reading frame of the Toledo genomic region. Detection of polynucleotide sequences can be by any suitable method, including but not limited to PCR amplification using suitable primers, LCR, hybridization of a labeled polynucleotide probe, and the like. Detection of
30 polypeptide species is typically done by immunoassay using a specific antibody to the Toledo region gene product(s).

The invention also provides a method of treating or preventing CMV infection, the method comprising administering to an individual an efficacious dose of a polypeptide which is
35 substantially identical to the deduced amino acid sequence of UL148. In a variation, the polypeptide is a truncated variant, mutein, or analog of the deduced amino acid sequence of UL148,

EXPERIMENTAL EXAMPLES

Overview

5 The growth advantage of Toledo in the SCID-hu mouse model resides in the genetic information encoded by the additional sequences (Toledo genomic region) we have identified. One gene in particular, UL148, has been mutagenized in Toledo by insertion of a selectable marker (*gptILacZ*) and the Toledo-based recombinant has been shown to replicate less well than Toledo in the SCID-hu assay. The genetic information of the corresponding
10 region of the avirulent Towne virus has been deduced by nucleotide sequence analysis and demonstrated to lack an open reading frame in Towne. UL148 can be considered to be representative of a "virulence determinant" for Toledo. The new Toledo sequence identified at the inverted repeats has been analyzed to reveal novel genes in Toledo. Deletion of genes encompassing UL138 to UL148 in recombinant viruses have been tested for growth properties in the SCID-hu (thy/liv) mouse. These recombinants have been shown to replicate to levels similar to the Towne virus and represent attenuated vaccine candidates,
15 since Towne has been shown to be safe and avirulent in humans. Such recombinants should show increased immunogenicity owing to their greater similarity to low passage virulent strains over that shown by highly-passaged Towne in humans. In addition, these strains should not exhibit the fully virulent phenotype shown by unmodified Toledo in humans due to the alterations we have introduced into their genomes.

This invention describes new recombinant HCMV viruses not previously described which are attenuated in virulence relative to low passage, virulent isolates by virtue of deletion
20 of sequences shown to be present in low passage, virulent isolates but which are lacking in laboratory strains. The identification of these sequences was essential in order to prepare transfer vectors capable of shuttling deletions (or insertions such as selectable markers) resulting in an effective
25 removal of coding information. Knowledge of the ORF usage on these DNAs permits deletion or insertion of one DNA into the other to specifically disrupt existing coding information. In

addition, this invention identifies sequences which can be used as "**spacer**" DNA for substitution into deleted regions of HCMV clinical isolates for purposes of attenuation.

5 Cosmid subclones of Towne and Toledo

Cosmid subclones of the CMV(Towne) and CMV(Toledo) genomes were constructed according to the method of Kemble et al. (1996) J. Virol. 70: 2044, incorporated herein by reference. Human foreskin fibroblast (HF) cells were infected with either Towne or Toledo and following the development of extensive CPE, DNA was isolated from nucleocapsids by a procedure similar to that used for the preparation of HSV nucleocapsids (Denniston et al. (1981) Gene 15: 365, incorporated herein by reference). The DNA was partially digested with *Sau3AI*, fractionated by agarose gel electrophoresis, and ligated to the *Bam*HI site of *Bam*HI, *Xba*I digested arms of the SuperCos•A1 cosmid vector. SuperCos•A1 was derived from SuperCos-1 (Stratagene, San Diego, CA) by the insertion of an oligonucleotide incorporating *Srf*I and *Pac*I recognition sequences flanking a unique *Bam*HI site. The position of the cosmid subclones relative to the viral genome was identified by Southern and DNA sequence analyses.

Overlapping cosmids for virus regeneration

Mapping the extent of the viral insert within the cosmid subclones was used as a basis to form specific Towne/Toledo chimeric viruses by choosing the appropriate cosmids from each virus. The ends of adjacent cosmids should overlap (~200bp or more) such that homologous recombination is permitted in eucaryotic cells.

To construct a Toledo based virus which lacked the Toledo U_L/b' region and in its place contained the Towne U_L/b' region, the following set of cosmids was used: Tol29, Tol158, Tol182, Tol122, Tol158, Tol24, Tn39, and Tn50. The resulting virus was designated Tol/Twn 39/50 (see Fig. 5). Other viruses were regenerated by cosmid cotransfection which lacked portions of the Toledo U_L/b' region. Toledo based viruses were generated by the cotransfection of the Toledo cosmids, Tol29, Tol158, Tol182, Tol122, Tol158, Tol24, Tol 212, Tol187 OR Tol159, Tol150,

Tol239, Tol235, Tol158, Tol24, Tol212, Tol187. Towne/Toledo chimeras lacking portions of the Toledo U_L/b' region were regenerated by cotransfection of Tn43, Tn13, Tn24, Tn9, Tn42, Tn51, Tol212, Tol187. Because Tol 212 and Tol187 did not overlap, deletions resulted in the viruses regenerated from these cosmid sets which lacked varying portions of the Toledo U_L/b' region.

Preparation of cosmids for cotransfection

A set of overlapping cosmid clones constituting the appropriate viral genome were individually digested with PacI to release the intact viral insert from the cosmid vector. The restriction enzyme was inactivated by heating at 65°C for 20 minutes, the cosmids were combined and the DNA precipitated with ethanol. A CaPO₄ precipitate was formed from approximately 8 to 16µg of this mixture and transfected using general transfection methods. The DNA was transfected into approximately 1 X 10⁶ low passage (<15 passes) HF, LF (human embryonic lung fibroblast) or IFIE1.3 (a gift of Ed Mocarski; these cells are immortalized HF cells that express the CMV major immediate early protein) cells. All these cells are permissive for CMV replication.

For HF and LF cells, approximately 1 X 10⁶ cells were plated onto a 25cm² flask 3 to 5 hours prior to the addition of the DNA-CaPO₄ precipitate. At this point, the precipitate was adsorbed directly to the cell monolayer for 30 minutes prior to the addition of media. 2ml of media was added and incubation continued for 4 hours at 37°C.

For IFIE1.3 cells, the cells were trypsinized approximately 16 hours prior to the addition of the DNA-CaPO₄ precipitate and seeded at a 1:2 density. At the appropriate time post seeding, the DNA-CaPO₄ precipitate was added in addition to 2ml of media and incubated at 37°C for 4 hours.

Following the 4 hour incubation, the DNA-CaPO₄ precipitate was removed, the cells incubated at 37°C for 3 min in 15% glycerol in Hepes buffered saline, rinsed one time with media and fed with 5ml of media. The media on the cells was changed every 3 to 4 days and plaques appeared in 10 to 21 days.

Construction of recombinant CMV by insertion of a *gpt/LacZ* marker

Two plasmids encompassing the Toledo U_L/b' region and derivatives thereof were constructed which contained a marker gene. A segment of DNA encompassing AD169 bases 156251-174483 was removed from pON2601 (Cha et al. (1996) J. Virol. 70: 78, incorporated herein by reference) and a *PacI* linker was introduced at AD169 base 174484 to yield a subclone of pON2601. Figs. 3 and 4 show a schematic drawing of the open reading frames in the Toledo U_L/b' region using sequence numbering from the Toledo U_L/b' region DNA insert. A 4.8 kb DNA fragment containing the *E. coli gpt* and *lac Z* genes driven by the HSV thymidine kinase and β actin promoters (Prichard et al. (1996) J. Virol. 70: 3018, incorporated herein by reference), respectively, was then inserted into the *NsiI* site in Toledo UL148 within the pON2601 subclone. The resulting plasmid containing the *gpt* and *lacZ* insert in UL148 was designated pGD6. Toledo open reading frames UL138 to UL148 were removed from pGD6 by a *BamHI* collapse to produce the plasmid pGD7. Toledo recombinant viruses Tol pGD6 and Tol pGD7 were constructed using plasmids pGD6 and pGD7, respectively, as described (Prichard et al. (1996) op.cit.

Analysis of recombinant CMV in SCID-hu (thy/liv) mice

SCID-hu (thy/liv) mice were derived by implanting human fetal thymus and liver beneath the kidney capsule of a female C.B. -17 *scid/scid* IcrTac mouse (McCune et al. (1988) Science 241: 1632, incorporated herein by reference). The SCID-hu (thy/liv) mouse model serves as an animal model that can distinguish virulent from avirulent strains of CMV based on their replication levels within the human implant (Mocarski et al. (1993) Proc. Natl. Acad. Sci. (U.S.A.) 90: 104 and Brown et al. (1995) J. Infect. Dis. 171: 1599, each incorporated herein by reference). Several weeks following implantation, the human implant on the murine kidney was surgically exposed and an inoculum of $\sim 10^4$ PFU of the appropriate virus was injected directly into the human tissue in a volume of 10 - 25 μ l. The murine kidney/human implant was placed back into the animal in

its natural position and the animal was recovered. 2 weeks following infection of the human tissue, the animal was sacrificed and the implant was removed and added to 2ml of 4.5% skim milk/50% media.

5 The excised implant was homogenized with an automated Dounce apparatus (Glas-Col, Terre Haute, IN) and the suspension was stored at -80°C until the titers were determined. The suspension was thawed at 37°C, sonicated on ice by three cycles of 10 sec on/ 10 sec off and centrifuged to remove the debris.
10 The supernatant was recovered and the titer of CMV present was determined on confluent monolayers of HF cells. 7 to 10 days after plating the virus, the monolayers were fixed and stained with Giemsa and plaques enumerated.

15 Fig. 6 shows results from this experiment. The virulence of the Toledo strain CMV is attenuated by functional disruption of Toledo genome region open reading frames.

20 The difference in virulence between the Towne and Toledo strains appears to have resulted from genetic differences generated during the adaptation of Towne to growth in diploid fibroblasts in culture. Both Towne and Toledo were originally isolated from the urine of a congenitally infected infant. Towne was subsequently passaged over 125 times in culture resulting in genetic alterations in the viral genome and an avirulent virus.
25 The virulent Toledo virus, in contrast, was passaged approximately 5 times in diploid fibroblasts in order to produce material that could cause disease in humans.

30 These linked genetic and biological differences can be used to create a live, attenuated HCMV vaccine. The rationale for tissue culture adaptation of Towne was to generate a live, attenuated vaccine strain. Towne has been shown to be safe and somewhat immunogenic. Towne, however, is overattenuated. The immune response induced by inoculation with Towne does not protect against subsequent HCMV infection as effectively as that generated by natural infection. Vaccine candidates can be
35 generated by replacing genetic elements of the overattenuated Towne strain with homologous portions of the virulent Toledo strain. Through the analysis of these "chimeric" viruses, a

skilled artisan can select those that have the level of desirable characteristics of Towne and be attenuated to a more efficacious degree.

Our first four chimeric viruses, as a set, will contain the entire Toledo genome introduced into the Towne genetic background. Each individual chimera of the set will contain approximately 40-55% of the Toledo genome; the remainder will be derived from Towne. Each of these chimeras will contain the UL/b' region derived from Toledo. Genes within this region of the Toledo genome can affect cell tropism of HCMV. The viruses, designated chimera I, II, III, and IV were constructed from the cosmids listed in Table 1 (see also Figs. 9 and 10).

Table 1. Cosmids used to generate specific chimeras.

Viruses	I	II	III	IV
	Tn46a	Tn46	Tn46	Tol29
	Tol58b	Tn45	Tn45	Tol58
	Tol182	Tol239	Tn23	Tn23
	Tn47	Tol22	Tn47	Tn47
	Tn44	Tol158	Tol184	Tn44
	Tn26	Tn26	Tol24	Tn26
	Tn20	Tn20	Tol212	Tol212
	Tol11	Tol11	Tol11	Tol122

Large quantities of each cosmid were prepared by purification of the E.coli produced material over a Qiagen column as described by the manufacturer. 10 micrograms of each cosmid was digested with the restriction enzyme Pac I (New England Biolabs) to physically separate cosmid vector from viral sequences. Following digestion, the enzyme was inactivated by incubation at 65oC for 20 minutes and the appropriate cosmids were combined, precipitated with ethanol in the presence of 0.3M sodium acetate, rinsed in 70% ethanol, and air-dried briefly. The resulting DNA was solubilized in approximately 100

microliters of 10mM Tris pH 7.5/1mM EDTA. Various amounts of the cosmid mix was transfected by the calcium phosphate technique into permissive fibroblast cells, specifically human lung fibroblasts (LF, prepared in our laboratory), human neonatal foreskin fibroblast (HF, a gift of Dr. Ed Mocarski, Stanford University), MRC-5 (ATCC) or IFIE1.3 cells (a gift of Ed Mocarski, Stanford University). The IFIE1.3 cell line constitutively expresses the HCMV iel gene product and has been transformed with the human papilloma virus E6 and E7 genes transduced by a retrovirus vector. 3 to 5 hours after transfection the cells were shocked by incubation in 15% glycerol/Hepes buffered saline for 3 minutes at 37°C and fed with DME/10% fetal bovine serum. 7 to 10 days after transfection plaques with distinct HCMV CPE were evident.

Plaques derived from the chimeras were allowed to grow until 100% of the monolayer exhibited CPE. At this point, DNA was extracted from the supernatant and cellular fractions and analyzed by restriction enzyme digestion. The structures of the viruses can be deduced by the cosmids used for construction of the chimera and confirmed by comparing the EcoRI digestion pattern to the maps derived for Towne and Toledo (see Fig. 10). Table 2 describes the composition of each of the chimeras, the nucleotide limits are derived from sequence analysis of the end of each cosmid insert and its homology to the AD169 strain of HCMV, which has been sequenced in its entirety. All of the chimeras had restriction enzyme patterns consistent with the proposed structures.

Table 2. Genetic composition of the chimeras.

<u>Chimera</u>	<u>Towne DNA</u>	<u>Toledo DNA</u>	<u>Crossover Region</u>
I	1 - 3799	15750-67568	3800-15749
	81647-170499	175069-203136	67569-81646
	205803 to S term.		170500-175068

II	1 - 47985	53244 --110000	47986-53243
	138834 - 170499	175069 - 203136	~110000- 130833
	205803 to S term.		170500 - 175068
			203137-205802

5

III	1 - ~99000*	108094 - 203136	~99000-108093
	205803 to S term		203137-205802

10

IV	43981 - 145583	1 - 41356	41357 - 43980
		150754 - S term	145584 - 150753

* Sequence at end of cosmid was undefinable. Nucleotide number is not exact crossover region is the region of cosmid overlap. The contribution of each virus to this region has yet to be defined.

Two other chimeras are constructed based on an observation derived from the sequence analysis of several different members of the beta-herpesvirus family including HCMV, human herpesvirus 6 (the causative agent of roseola) and murine cytomegalovirus. Representative members of each of these viruses have been sequenced in their entirety and a "core" set of genes corresponding to HCMV UL23 to UL122 are conserved among these evolutionarily divergent entities (Chee, et al. 1990; Gompels, et al. 1995; Rawlinson, et al. 1996, incorporated herein by reference). These core genes contribute to DNA replication, virion structure, and other basic features of the virus. Genes outside this core region are involved in virus-host and virus-immune system interaction and may determine specific properties of virus biology. Replication of the Chimeras was tested in SCID-hu mice having a thy/liv sandwich under the kidney capsule; representative data is shown in Figure 11.

Two additional chimeras are constructed: one which has the core derived from Toledo with the remainder of the genes derived

from Towne and an inverse construct in which the core is derived from Towne and the remainder of the genes are from Toledo. These viruses are constructed through the use of overlapping cosmids and derivatives of the cosmids. Table 3 outlines

5 the constructs that can be used to generate these two chimeric viruses.

Table 3. Construction of chimeras containing conserved core regions.

<u>Towne Core/Tol Noncore</u>	<u>Tol Core/Towne Noncore</u>
Tol 29	Tn43
Tol158 nts: 3800 - 27862	Tn45 nts: 7854 - 27862
Tn45 nts: 27500 - 53243	Tol 58: 27500 - 43980
Tn23	Tol 182
Tn47	Tol 22
Tn44	Tol 158
Tn26	Tol 24
Tn20	Tol212 nts:145584-17200
Tol11nts: 170852 - 188890	Tn 39 nts:170852 - 183512
Tol122	Tn 15

All of these viruses are used to inoculate healthy adult human volunteers. These individuals are assessed for symptoms of HCMV disease, including fever, malaise, and abnormal liver enzyme levels. Hallmarks of viral infection are also assessed by measuring HCMV specific antibody titers before and after inoculation as well as viral culture for the isolation of infectious virus from bodily fluids. A successful vaccine candidate is identified as a strain that maintains the safety profile of Towne while stimulating a greater immune response to the virus.

Figure 12 shows schematic depiction of the Toledo genome in comparison with highly passaged Towne (short genome) and low-passage Towne (long genome).

The foregoing description of the preferred embodiments of the present invention has been presented for purposes of illustration and description. They are not intended to be
5 exhaustive or to limit the invention to the precise form disclosed, and many modifications and variations are possible in light of the above teaching.

Such modifications and variations which may be apparent to a person skilled in the art are intended to be within the
10 scope of this invention.

SECRET - 5267260

CLAIMS:

1. A method for attenuating a cytomegalovirus comprising functionally disrupting an open reading frame of a Toledo genome region or its homologs.

2. A method of claim 1 for attenuating a CMV strain or isolate containing an encoding polynucleotide sequence encoding a polypeptide which is at least 80 percent sequence identical to a polypeptide encoded by UL138, UL139, UL140, UL141, UL142, UL143, UL144, UL145, UL146, UL147, and/or UL148 of the Toledo genomic region; the method comprising functionally inactivating said encoding polynucleotide sequence to produce a Toledo region-attenuated CMV variant.

3. A chimeric CMV virus, comprising a genome having a plurality of polynucleotide sequences, linked in conventional phosphodiester linkage, wherein at least two of said polynucleotide sequences are derived from different clinical isolates or strains of CMV.

4. The chimeric CMV virus of claim 3, wherein said chimeric CMV virus comprises a genome having a plurality of polynucleotide sequences, linked in conventional phosphodiester linkage, wherein a first CMV genome sequence of at least 500 bp and less than a complete CMV genome length is at least 98 percent sequence identical to a first CMV isolate or strain, and at least one additional CMV sequence of at least 500 bp and less than a complete CMV genome length is at least 98 percent sequence identical to a second CMV isolate or strain which has a genome having a polynucleotide sequence of at least 500 bp which is less than 60 percent sequence identical to any portion of the genome of said first CMV isolate or strain and/or which is absent or substantially absent in the genome of said first CMV isolate or strain.

5. The chimeric CMV virus of claim 4, wherein the chimeric

CMV virus genome is at least 80 percent sequence identical to a high-passage Towne genome and said chimeric CMV virus genome typically further comprises genetic information which is at least 80 percent sequence identical to at least 1 kbp of a virulence region of a clinical isolate of CMV or a low-passage strain of CMV other than low-passage Towne.

6. A chimeric CMV virus of claim 5, wherein a complete Toledo genome region is present.

7. A chimeric CMV virus of claim 3 which is Chimera I.

8. A chimeric CMV virus of claim 3 which is Chimera II.

9. A chimeric CMV virus of claim 3 which is Chimera III.

10. A chimeric CMV virus of claim 3 which is Chimera IV.

11. A chimeric CMV virus of claim 3 which is Towne/Toll1.

12. A method for producing an attenuated CMV virus, comprising forming a chimeric CMV virus having a genome comprising a first genome portion derived from a CMV clinical isolate or CMV strain having a virulence region and a second genome portion derived from a highly passaged CMV strain or isolate lacking a virulence region.

General Information		Performance		Energy Efficiency		Reliability		Scalability		Security		Compliance		Interoperability	
Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
System Name	Project Phoenix	Throughput (ops/sec)	1200	Power Consumption (W)	150	Uptime (hrs)	99.99	Latency (ms)	50	Encryption Type	AES-256	GDPR Compliant	Yes	API Version	2.1.0
Version	1.0.0	Latency (ms)	50	Efficiency (%)	85	Availability (%)	99.99	Throughput (ops/sec)	1200	Authentication Method	OAuth2	ISO 27001	Yes	Protocol Support	REST, GraphQL
Author	John Doe	Efficiency (%)	85	Reliability Score	9.5	Scalability Factor	10x	Security Audit Date	2023-10-27	Logging Level	Verbose	PCI DSS	Yes	Integration Points	15
Reviewer	Jane Smith	Reliability Score	9.5	Scalability Factor	10x	Security Audit Date	2023-10-27	Logging Level	Verbose	Integration Points	15	Compliance Frameworks	GDPR, HIPAA	Interoperability Score	9.0
Project Lead	Michael Chen	Security Audit Date	2023-10-27	Logging Level	Verbose	Integration Points	15	Compliance Frameworks	GDPR, HIPAA	Interoperability Score	9.0	Documentation Status	Complete	Deployment Status	Staging
QA Lead	Sarah Johnson	Documentation Status	Complete	Deployment Status	Staging	Support Channels	24/7	Feedback Mechanism	Survey	Release Date	2023-11-01	Next Review Date	2024-01-01	Project Status	On Track
Dev Lead	David Kim	Support Channels	24/7	Feedback Mechanism	Survey	Release Date	2023-11-01	Next Review Date	2024-01-01	Project Status	On Track	Overall Rating	A-	Recommendation	Proceed
Product Manager	Emily White	Release Date	2023-11-01	Next Review Date	2024-01-01	Project Status	On Track	Overall Rating	A-	Recommendation	Proceed	Notes	All metrics met or exceeded expectations. Minor improvements suggested for scalability.		

5

Toledo UL134

10	20	30	40	50	60
MARTREASPV	PPRSPMPHI	HTMIFSPAWN	LKLRVGKGR	TDIYALDFWK	RHFLARNVFI
70	80	90	100	110	120
VQTLRKEMCA	KSENSLSHRG	RVTFRSDAAA	VVVEPRPRPP	ARQLVPPRPR	RVASAAWRGE
130	140	150	160	170	180
ARRADRRALP	SAATVVVNSP	SVRTEVCLSV	YPSVYLSPYL	SSVWVPMSVL	AAAVG*....

Toledo UL135

10	20	30	40	50	60
MSVHRPFPTR	SLRFQAGEKI	MVWIWLGIGL	LGGTGLASLV	LAISLFTQRR	GRKRSDETSS
70	80	90	100	110	120
RGRLPGAASD	KRGACACCYR	NPKEDVVEPL	DLELGLMRVD	THPPTPQVPR	CTSLYIGEDG
130	140	150	160	170	180
LPIDKPEFPP	ARFEIPDVST	PGTPTSIGRS	PSHCSSSSSL	SSSTSVDTVL	YQPPPSWKPP
190	200	210	220	230	240
PPPGRKKRPP	TPPVRAPTTR	LSSHRPPTPI	PAPRKNLSTP	PTKKTTPPTK	PKPVGWTPPV
250	260	270	280	290	300
TPRPFKPTPT	PQKPPRNPRL	PRTVGLENLS	KVGLSCPCPR	PRTPTPTTL	PIVSVSELAP
310	320	330	340	350	360
PPRWSDIEEL	LEQAVQSVMK	DAESMQMT*

Toledo UL136

10	20	30	40	50	60
MSVKGEMPE	MTWDLVRNK	WRRRKALSRI	HRFWECLRV	WWLSDAGVRE	TDPPRPRRRP
70	80	90	100	110	120
TWMTAVFHVI	CAVLLTLMIM	AIGALIAYLR	YYHQDSWRDM	LHDLFCGCHY	PEKCRHHER
130	140	150	160	170	180
QRRRRQAMDV	PDPELGDPAR	RPLNGAMYYG	SGCRFDTVEM	VDETRPAPPA	LSSPETGDDS
190	200	210	220	230	240
NDDAVAGGGA	GGVTSPATRT	TSPNALLPEW	MDAVHVAVQA	AVQATVQVSG	PRENAVSPAT
250	260	270	280	290	300
*

Fig. 2B

Toledo UL130

10 20 30 40 50 60
MLRLLLRHHF HCLLLCAVWA TPCLASPWST LTANQNPSPP WSKLTYSKPH DAATFYCPFL
70 80 90 100 110 120
YPSPPRSPLQ FSGFQQVSTG PECRNETLYL LYNREGQTLV ERSSTWVKKV IWYLSGRNQT
130 140 150 160 170 180
ILQRMPTAS KPSDGNVQIS VEDAKIFGAH MVPKQTKLLR FVVNDGTRYQ MCVMKLESWA
190 200 210 220 230 240
HVFRDYSVSF QVRLTFTEAN NQTYTFCTHP NLII*.....

Toledo UL132

10 20 30 40 50 60
MPALRGPLRA TFLALVAFGL LLQIDLSDAT NVTSSSTKVPT STSNRNMVDN ATSSGPTTGI
70 80 90 100 110 120
NMTTTHESV HNVRNNEIMK VLAILFYIVT GTSIFSFIIV LIADVYSSCC KHPGRFRFAD
130 140 150 160 170 180
EEAVNLLDDT DDSGGSSPFG SGSRRGSQIP PDFVPRALIS GWKLGTTGTRR RRRPRPASA*
190 200 210 220 230 240
NMILRTSSIS ERMATWTRRS *IPIMGEARL *PSNLTSRTM RRTPSGTTFR CTMN*PPRKW
250 260 270 280 290 300
KNLRTAPAGR FPN**KLPCN PSRSEIPSTT R.....

Toledo UL133

10 20 30 40 50 60
MGCDVHDPSW QCQGWVPTII VAWITCAALG IWCLAGSSAD VSSGPGIAAV VGCSVFMIFL
70 80 90 100 110 120
CAYLIRYREF FKDSVIDLLT CRWVRYCSCS CKCSCKCISG PCSRCCSACY KETMIYDMVQ
130 140 150 160 170 180
YGHRRRPGHG DDPDRVICEI VESPPVSAPT VSVPPPSEES HQPVIPPQPP APTSEPKPKK
190 200 210 220 230 240
GRAKDKPKGR PKDKPPCEPT VSSQPPSQPT AMPGGPPDAP PPAMPQMPPG VAEAVQAAVQ
250 260 270 280 290 300
AAVAAALQQQ QQHQTGT*..

Fig. 2A

Toledo UL137

10 20 30 40 50 60
MATISTSITP MMGNPTFSGR SSMVTVLCPD LRPSLSLLYS TRAGTAPSTL LRSGRYGVLP
70 80 90 100 110 120
RATYLHGRLN GGLDRMHRI HPFWQOCVRR RRTSRG*... ..

Toledo UL138

10 20 30 40 50 60
MDDLPLNVGL PIIGVMLVLI VAILCYLAYH WHDTFKLVLM FLSYRWLIRC CELYGEYERR
70 80 90 100 110 120
FADLSSLGLG AVRRESDRRY RFSERPDEIL VRWEEVSSQC SYASSRITDR RVGSSSSSSV
130 140 150 160 170 180
HVASQRNSVP PPDMAVTAPL TDVDLLKPVT GSATQFTTVA MVHYHQEYT*

Toledo UL139

10 20 30 40 50 60
MLWILVLFAL AASASETTTG TSSNSSQSTS ATANTTVSTC INASNGSSWT VPQLALLAAS
70 80 90 100 110 120
GWTLSGLLLL FTCCFCCFWL VRKICSCCGN SSESESKTTH AYTNAFTSS DATLPMGTTG
130 140 150 160 170 180
SYTPPDGGSF PPPPR*.... ..

Toledo UL140

10 20 30 40 50 60
MTPAQTNATT TVHPHDAKNG SGGALPTLV VFGFIVTLLF FLFMLYFWNN DVFRKLLRAL
70 80 90 100 110 120
GSSAVATAST RGKTRSSTVV HHVVPRATTR VVLTACHRTF FYHPRPMAVL TTRH*.....

Toledo UL141

10	20	30	40	50	60
MRQVAYRRRR	ESSCAVLVHH	VGRDGDGE	AAKKTCKKTG	RSVAGIPGEK	LRRTVVTTTP
70	80	90	100	110	120
ARRLSGRHTE	QEQAGMRLCE	KGKKRIIMCR	RESLRTLPLW	FWVLLSCPRL	LEYSSSSFPF
130	140	150	160	170	180
ATADIAEKMW	AENYETTSPA	PVLVAEGEQV	TIPCTVMTHS	WPMVSIRARF	CRSHDGSDEL
190	200	210	220	230	240
ILDAVKGHRL	MNGLQYRLPY	ATWNFSQLHL	GQIFSLTFNV	SMDTAGMYEC	VLRNYSHGLI
250	260	270	280	290	300
MQRVILTQL	ETLSRPDEPC	CTPALGRYSL	GDQIWSPTPW	RLRNHDCGTY	RGFQRNYFYI
310	320	330	340	350	360
GRADAEDCWK	PACPDEEPDR	CWTVIQRYRL	PGDCYRSQPH	PPKFLPVTTP	PPADIDTGMS
370	380	390	400	410	420
PWATRGIAAF	LGFSWIFTVC	FLCYLCYLQC	CGRWCPTPGR	GRRGGEGYRR	LPTYDSYPGV
430	440	450	460	470	480
RKMKR*....

Toledo UL142

10	20	30	40	50	60
MRIEWVWWLF	GYFVSSVGSE	RSLSYRYHLE	SNSSTNVVCN	GNISVFNVTG	LGVRYNITVG
70	80	90	100	110	120
ISSSLIGHL	TIQVLESWFT	PWVQNKSYNK	QPLGDTETLY	NIDSENIHRV	SOYFHTRWIK
130	140	150	160	170	180
SLQENHTCDL	TNSTPTYTYQ	VNVNNTNYLT	LTSSGWQDRL	NYTVINSTHF	NLTESNITSI
190	200	210	220	230	240
QKYLNTTCIE	RLRNYTLESV	YTTTVPQNIT	TSQHATTTMH	TIPPNTITIQ	NTTQSHTVQT
250	260	270	280	290	300
PSFNDTHNVT	KHTLNISYVL	SQKTNNTTSP	WIYAIPMGAT	ATIGAGLYIG	KHFTPVKFBVY
310	320	330	340	350	360
EVWRGQ*...

Fig. 2D

Toledo UL143

10 20 30 40 50 60
MARSVKTIRI QHIYSRSSN TLQHMSKKQE SIATITFGRI TCCHPLASIN LMFNGSCTVT
70 80 90 100 110 120
VKISMGINGS TNVHQLVIVL HLGNNRCQPWR QV*.....

Toledo UL144

10 20 30 40 50 60
MKPLIMLICF AVILLQLGVT KVCQHNEVQL GNECCPPCGS GQRVTKVCTD YTSVTCTPCP
70 80 90 100 110 120
NGTYVSGLYN CTDCTQCNTV QVMIRNCTST NNTVCAPKNH TYFSTPGVQH HKQRQQNHTA
130 140 150 160 170 180
HITVKQGKSG RHTLAWLSLF IFLVGIILLI LYLIAAYRSE RCQQCCSIGK IFYRTL*...

Toledo UL145

10 20 30 40 50 60
MCTDPRRTAG WERLTHASY HANYGAYAVL MATSQRKSLV LHRYSAVTAV ALQLMPVEIV
70 80 90 100 110 120
RKLDQSDWVR GAWIVSETFP TSDPKGVWSD DDSSMGGSD *.....

Toledo UL146

10 20 30 40 50 60
MRLIFGALII FLAYVYHYEV NGTELRCRCL HRKWPPNKII LGNYWLHRDP RGP GCDKNEH
70 80 90 100 110 120
LLYPDGRKPP GPGVCLSPDH LFSKWLDKHN DNRWYNVNIT KSPGPRRINI TLIGVRG*..

Toledo UL147

10	20	30	40	50	60
MVLTWLHHPV	SNSHINLLSV	RHLSLIAYML	LTICPLAVHV	LELEDYDRRC	RCNNQILLNT
70	80	90	100	110	120
LPVGTELLKP	IAASESCNRQ	EVLAILKDKG	TKCLNPNAQA	VRRHINRLFF	RLILDEEQRI
130	140	150	160	170	180
YDVVSTNIEF	GAWPVPTAYK	AFLWKYAKRL	NYHHFRLRW*

Toledo UL148

10	20	30	40	50	60
MLRLLFTLVL	LALHGQSVGA	SRDYVHVRL	SYRGDPLVFK	HTFSGVRRPF	TELGWAACRD
70	80	90	100	110	120
WDSMHCTPFW	STDLEQMTDS	VRRYSTVSPG	KEVTLQLHGN	QTVQPSFLSF	TCRLQLEPVV
130	140	150	160	170	180
ENVGLYVAYV	VNDGERPQQF	FTPQVDVVR	ALYLETLSRI	VEPLESGRLA	VEFDTPDLAL
190	200	210	220	230	240
APDLVSSLFV	AGHGETDFYM	NWTLRRSQTH	YLEEMALQVE	ILKPRGVRHR	AIHHHPKLQP
250	260	270	280	290	300
GVGLWIDFCV	YRYNARLTRG	YVRYTLSPKA	RLPAKAEGWL	VSLDRFIVQY	LNTLLITMMA
310	320	330	340	350	360
AIWARVLITY	LVSRRR*...

Toledo UL149

10	20	30	40	50	60
MVDQCCYRHL	HRSLSGGPDV	LYAAAGTQRE	QQRDKSLAA	TAPSAVAGPP	ADRDVVDHRT *
70	80	90	100	110	120
ETHAYETPRY	ATRCLETYTT	PVRSVVRRTT	CGKRVASQSP	PRSCLVAPQS	SPAHPPRHPE
130	140	150	160	170	180
GG*.....

Toledo UL150

10	20	30	40	50	60
MQLCSHSSISS	QRHVASSMHC	RSRHQRTPPS	ATTHGPCAPT	SRILRRLLTT	RRFLPRTSPS
70	80	90	100	110	120
SNTVCCIRRR	LHERTIRHSM	RCRRRDMASS	ASTPVSHTQP	LAANHRRSRI	TYATTDPTNS
130	140	150	160	170	180
PTASPAKSDK	LEADADPALH	RRPASLLRHL	FQPCHAQRGT	SNRATSQRAS	LNAVHHKLCG
190	200	210	220	230	240
AMISSSCSTT	CTPLIMDLPS	LSVELSAGHK	KKETPTEGGW	GGEEGEDDVL	ATIRNTLSAP
250	260	270	280	290	300
TSPAAATTHR	LSFPGESTFC	LTAVSECSQR	RTSTAALTPP	PPAVAAAFSF	SSTVSETGTF
310	320	330	340	350	360
PQSTTGRTRV	DDTAVVTAGD	PRSPVTHVTL	LQIFRLRSSL	LTSRSGGALR	GGEHEAIPKV
370	380	390	400	410	420
ASLFWTLLKA	TQIVEMTHKT	PSADSHRNPQ	KYTDRPQRL	LTALAIWQRT	YNDTRAAHAP
430	440	450	460	470	480
QVRLGDLT	YRRPQTATAS	TKAHTQQQPE	EPKGQOIWTQ	TAGQAAPHGD	EPHSDGELRR
490	500	510	520	530	540
ESHSAPPTSR	TLPDTILAVK	RRSVAQRSHV	RLDAKPGLNE	RDGFRQRLLL	PLSGYFRANE
550	560	570	580	590	600
LRNQQFMGYG	TKNGLKNTWL	TRPLGVAGGV	RETIGERQDR	NVADSATQRV	FHTLYAALQT
610	620	630	640	650	660
VRVWYTALGT	AWRTSGSRTR	ESLFDGPRRR	DRQAARLRL	EL*.....

Fig. 2G

Variable	Mean	SD	Min	Max
Age	34.5	10.2	21	55
Gender	Male	Female		
Marital status	Married	Single		
Education	High school	College		
Occupation	Manager	Worker		
Income	\$10,000	\$20,000		
Health status	Good	Fair		
Exercise frequency	Weekly	Monthly		
Stress level	Low	High		
Smoking status	Smoker	Non-smoker		
Alcohol consumption	Regular	Occasional		
Family size	2	3		
Home ownership	Owner	Renter		
Commute time	15 min	30 min		
Work hours	40 hrs	50 hrs		
Job satisfaction	High	Low		
Life satisfaction	High	Low		
Overall health	Good	Fair		
Physical activity	Active	Sedentary		
Mental health	Good	Fair		
Emotional stability	Stable	Unstable		
Resilience	High	Low		
Optimism	Optimistic	Pessimistic		
Self-esteem	High	Low		
Empathy	High	Low		
Conscientiousness	High	Low		
Openness	High	Low		
Agreeableness	High	Low		
Neuroticism	Low	High		
Extraversion	High	Low		
Introversion	High	Low		
Shyness	High	Low		
Assertiveness	High	Low		
Leadership	High	Low		
Teamwork	High	Low		
Communication	High	Low		
Problem-solving	High	Low		
Decision-making	High	Low		
Time management	High	Low		
Organization	High	Low		
Productivity	High	Low		
Efficiency	High	Low		
Quality of work	High	Low		
Job performance	High	Low		
Work-life balance	High	Low		
Work satisfaction	High	Low		
Life satisfaction	High	Low		
Overall health	Good	Fair		
Physical activity	Active	Sedentary		
Mental health	Good	Fair		
Emotional stability	Stable	Unstable		
Resilience	High	Low		
Optimism	Optimistic	Pessimistic		
Self-esteem	High	Low		
Empathy	High	Low		
Conscientiousness	High	Low		
Openness	High	Low		
Agreeableness	High	Low		
Neuroticism	Low	High		
Extraversion	High	Low		
Introversion	High	Low		
Shyness	High	Low		
Assertiveness	High	Low		
Leadership	High	Low		
Teamwork	High	Low		
Communication	High	Low		
Problem-solving	High	Low		
Decision-making	High	Low		
Time management	High	Low		
Organization	High	Low		
Productivity	High	Low		
Efficiency	High	Low		
Quality of work	High	Low		
Job performance	High	Low		
Work-life balance	High	Low		
Work satisfaction	High	Low		
Life satisfaction	High	Low		
Overall health	Good	Fair		
Physical activity	Active	Sedentary		
Mental health	Good	Fair		
Emotional stability	Stable	Unstable		
Resilience	High	Low		
Optimism	Optimistic	Pessimistic		
Self-esteem	High	Low		
Empathy	High	Low		
Conscientiousness	High	Low		
Openness	High	Low		
Agreeableness	High	Low		
Neuroticism	Low	High		
Extraversion	High	Low		
Introversion	High	Low		
Shyness	High	Low		
Assertiveness	High	Low		
Leadership	High	Low		
Teamwork	High	Low		
Communication	High	Low		
Problem-solving	High	Low		
Decision-making	High	Low		
Time management	High	Low		
Organization	High	Low		
Productivity	High	Low		
Efficiency				

22

4

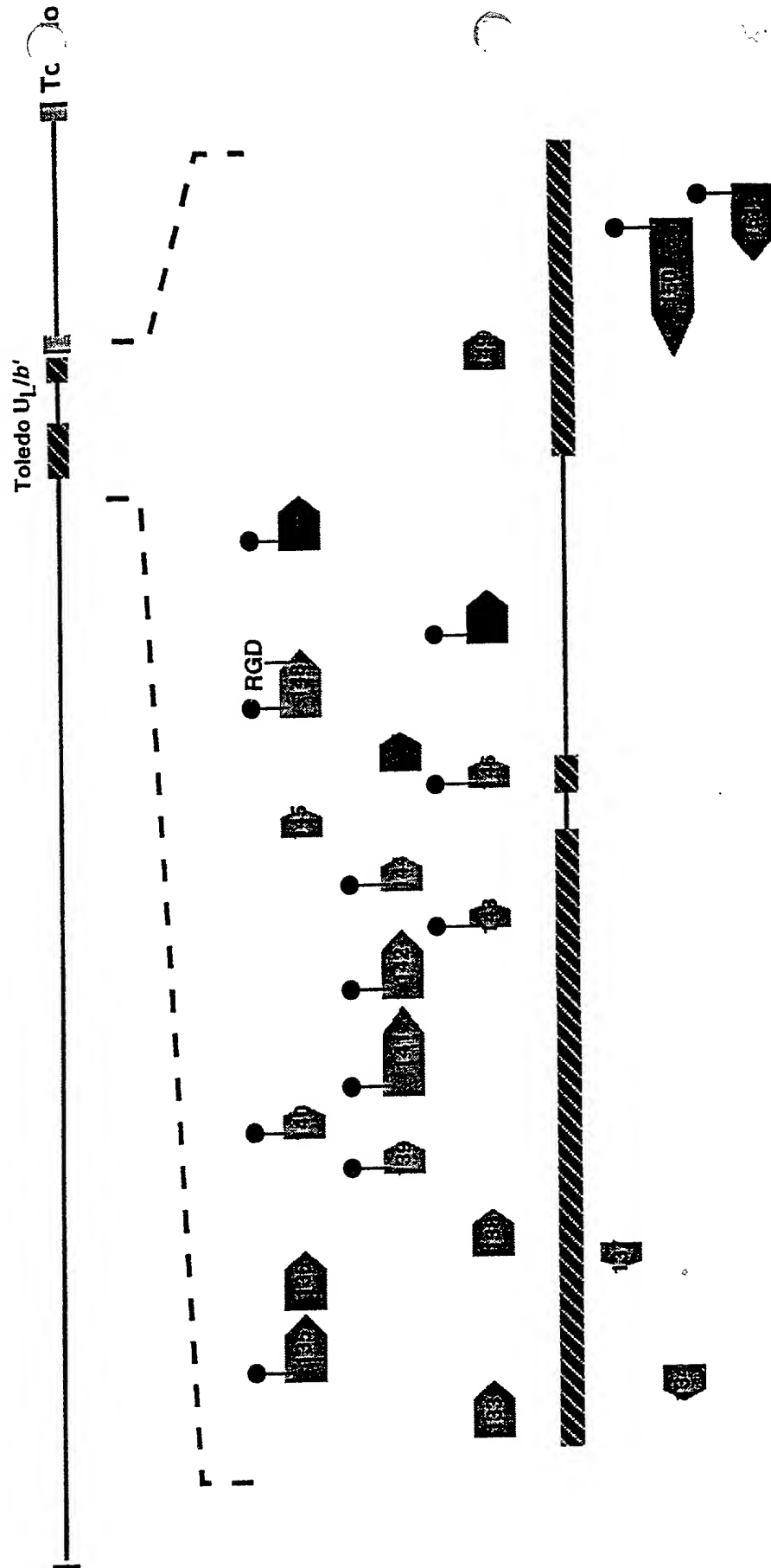
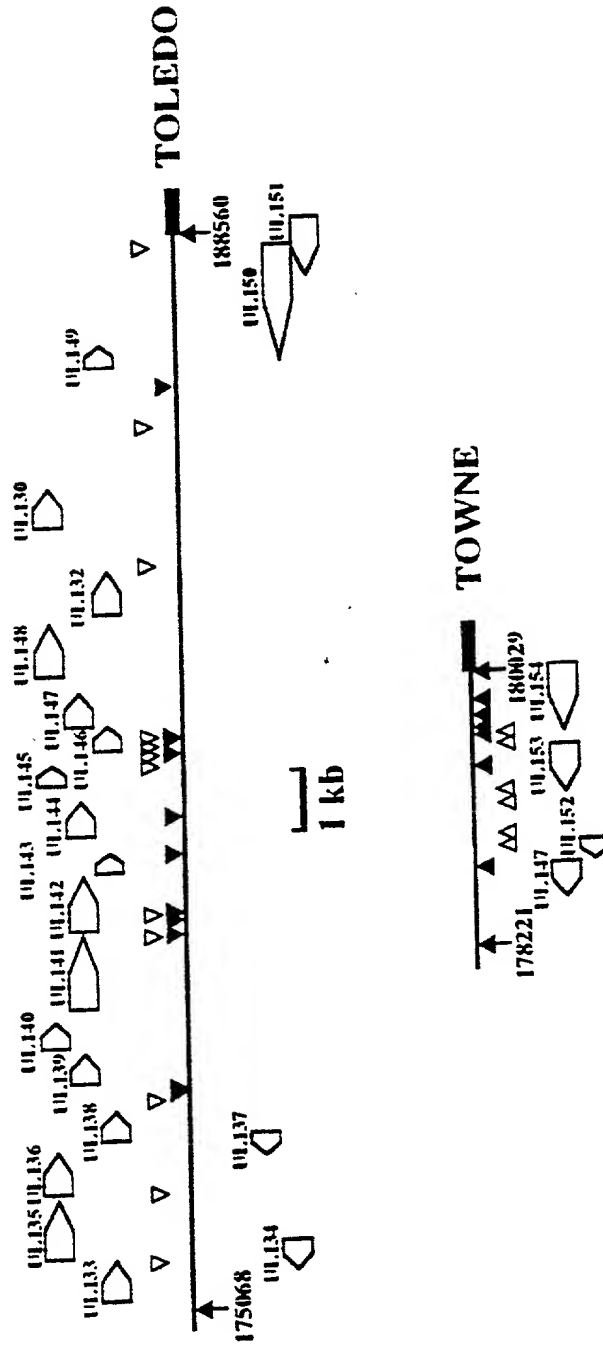
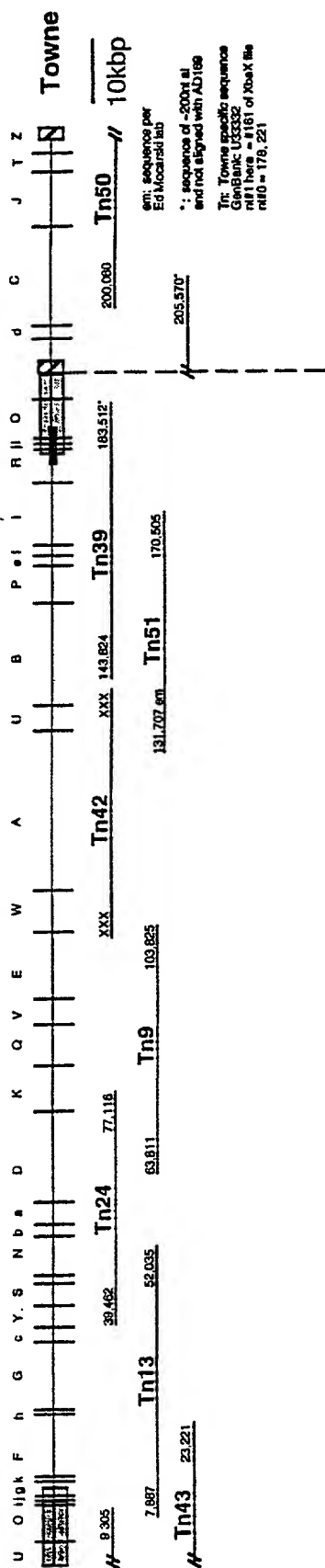


Fig. 3





em; sequence per
Ed Mocarid lab

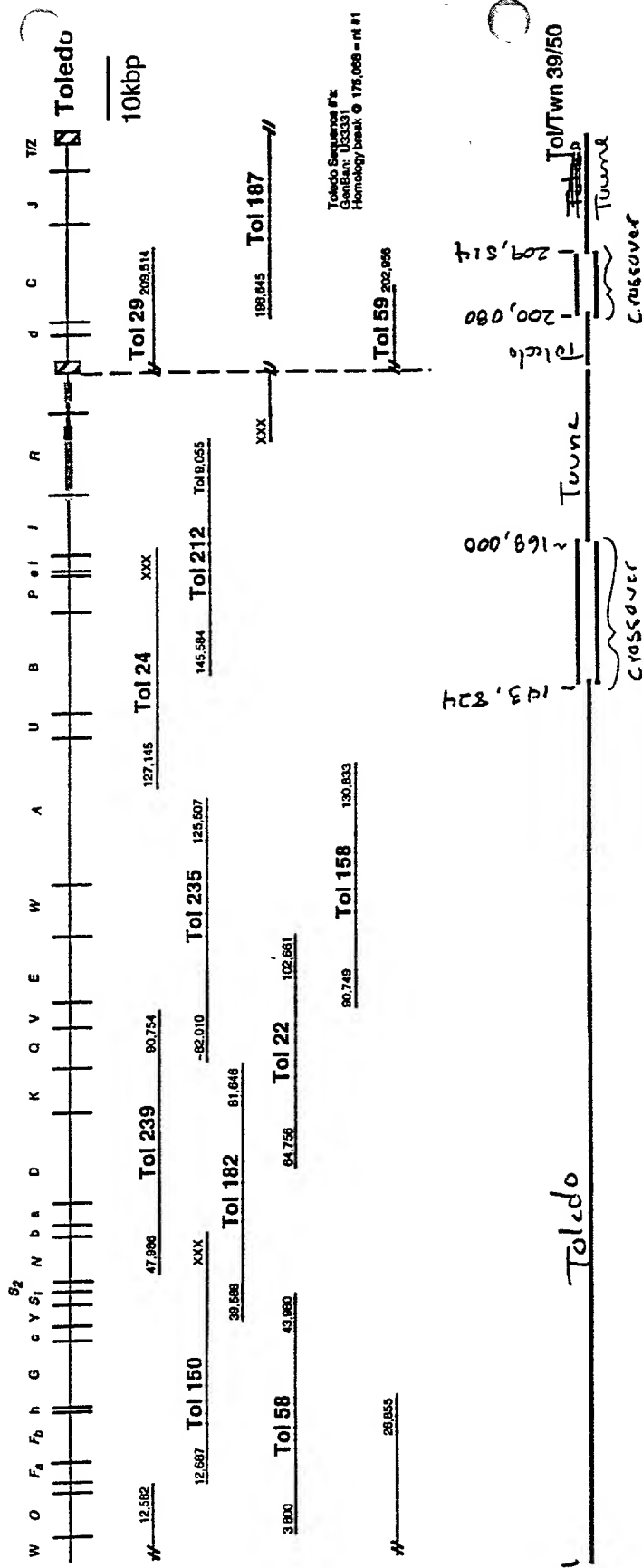


Fig. 5

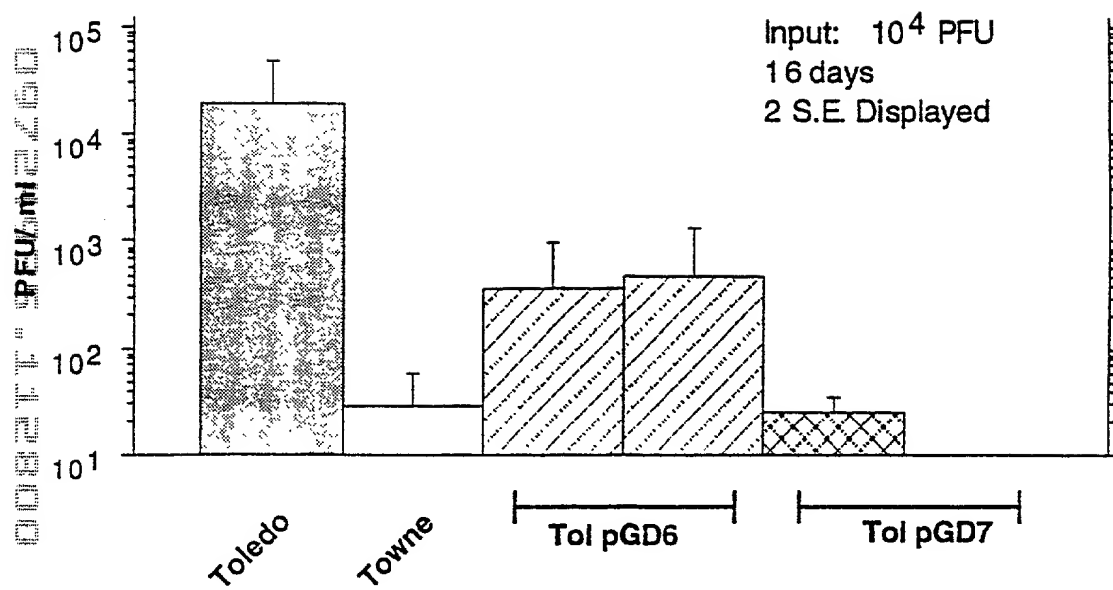


Fig. 6

000001 58542460

Clinical Strains of CMV Contain Sequences Homologous to the Toledo U_L/b' Region

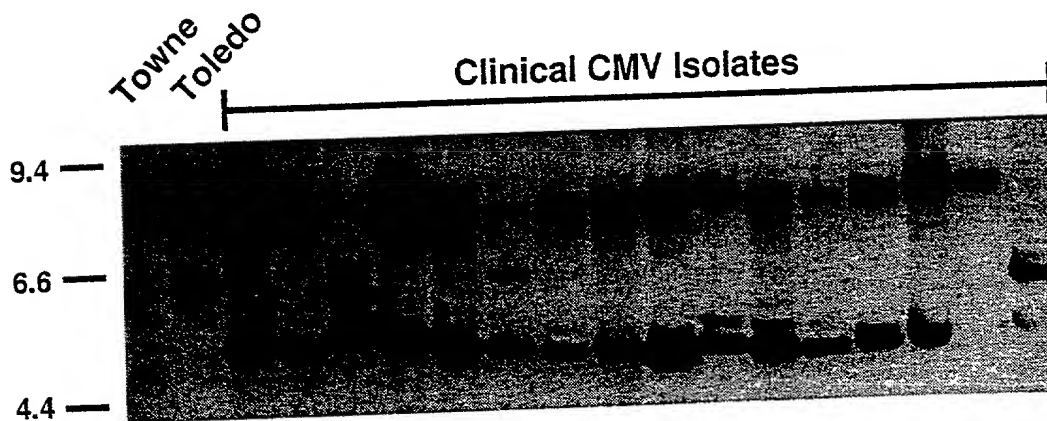


Figure 7

Previous Towne Vaccine Strains Hybridize to the Toledo U_L/b' Region

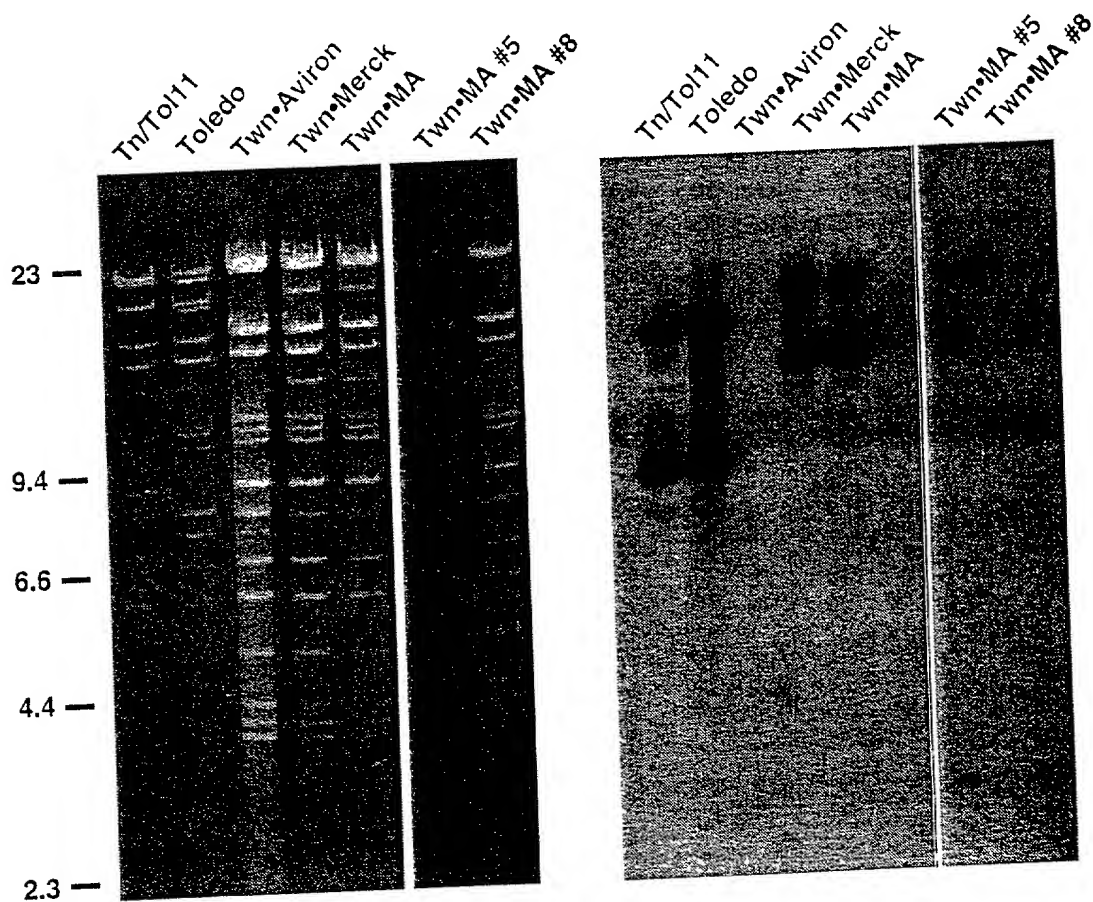
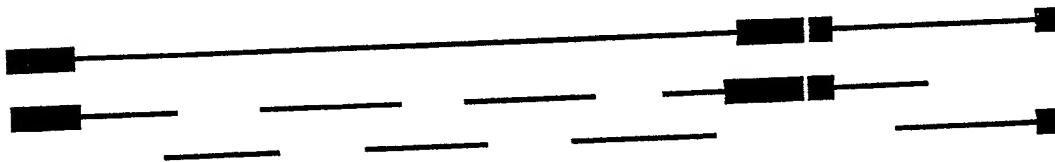


Figure 5

Cotransfection of Cosmids Regenerates Infectious CMV

Towne•AV



Toledo

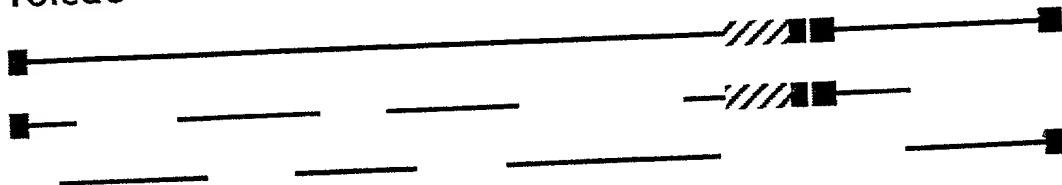


Figure 9

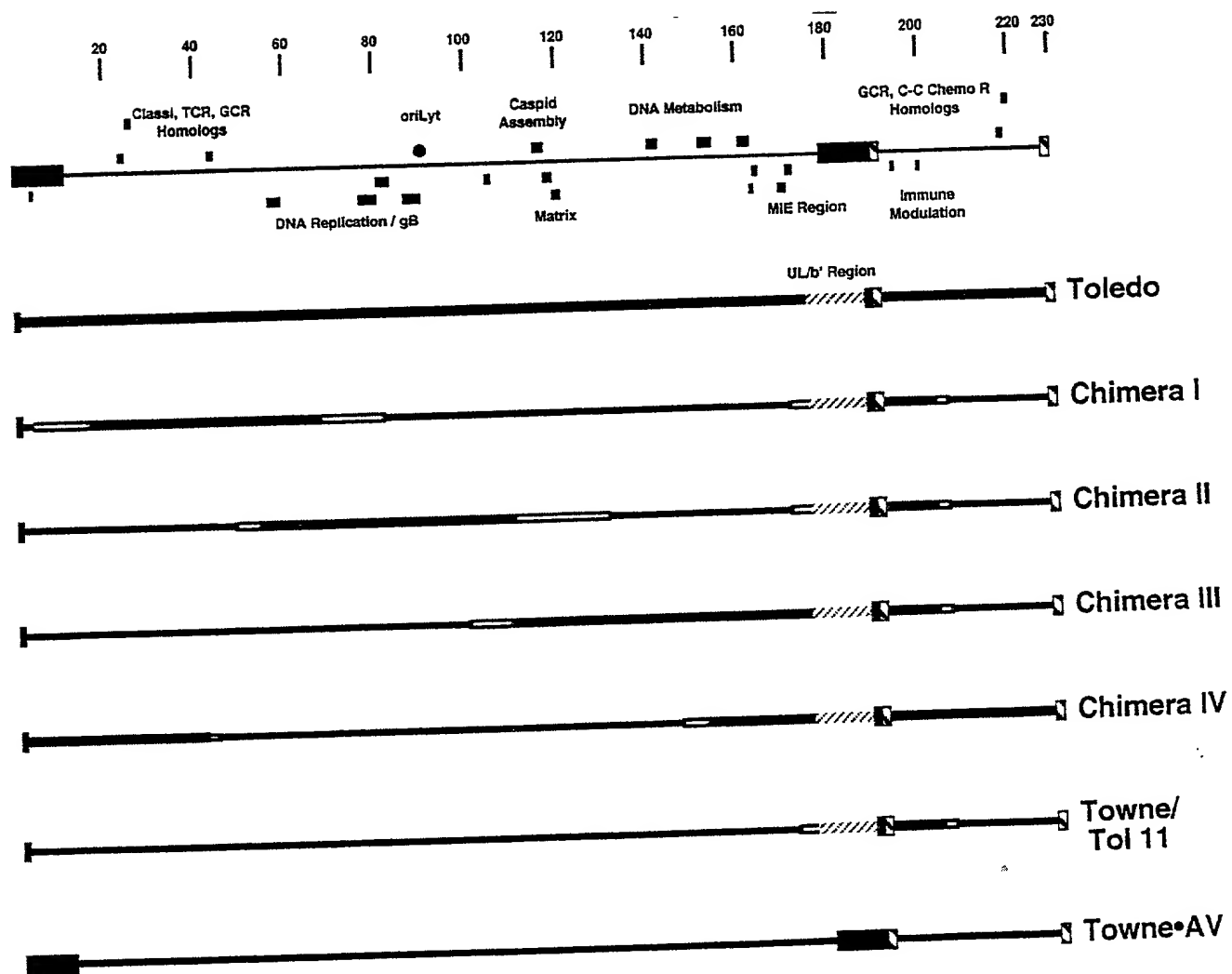


Figure 10

The Chimeras Replicate in the SCID-hu(thy/liv) Implant

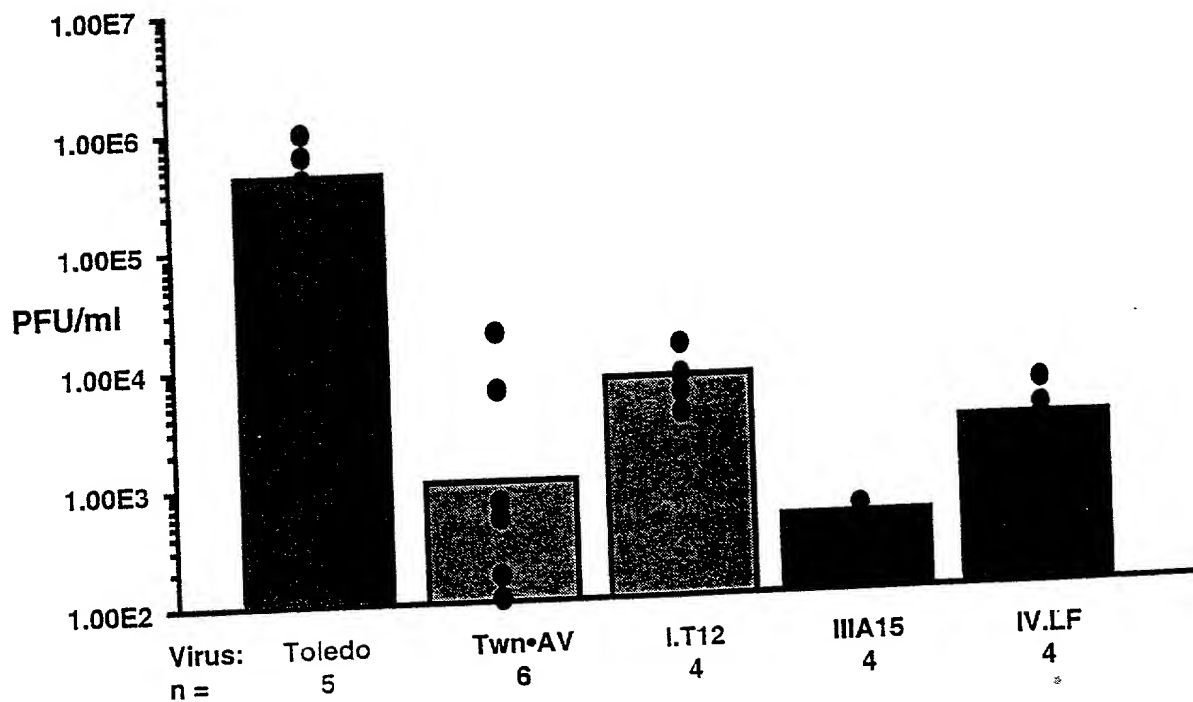


Figure 11

Comparison of the Towne (long) and Towne (short) Genotypes

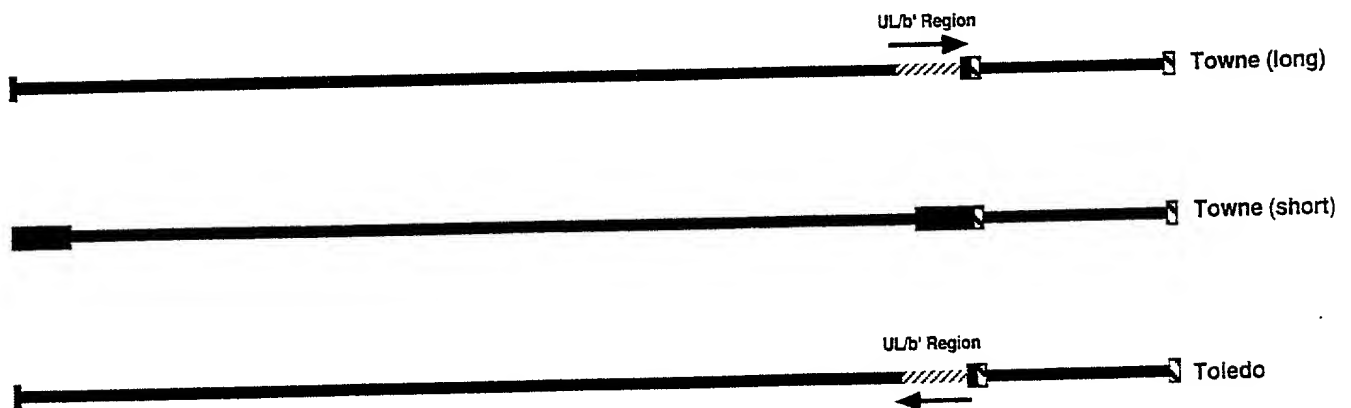


Figure 12

10	20	30	40	50	60
CGCTGTAGGG	ATAAATAGTG	CGATGGCGTT	TGTGGGAGAA	CGCAGTAGCG	ATGGGTTGCG
GCGACATCCC	TATTTATCAC	GCTACCGCAA	ACACCCTCTT	GCGTCATCGC	TACCCAACGC
70	80	90	100	110	120
ACGTGCACGA	TCCTTCGTGG	CAATGCCAAT	GGGGCGTTCC	CACGATTATC	GTGGCCTGGA
TGCACGTGCT	AGGAAGCACC	GTTACGGTTA	CCCCGCAAGG	GTGCTAATAG	CACCGGACCT
130	140	150	160	170	180
TAACATGCGC	GGCTTTAGGA	ATTTGGTGTT	TGGCGGGATC	GTGGGCGGAT	GTCTCTTCGG
ATTGTACGCG	CCGAAATCCT	TAAACCACAA	ACCGCCCTAG	CAGCCGCCTA	CAGAGAAGCC
190	200	210	220	230	240
GACCCGGCAT	CGCAGCCGTA	GTGGGCTGTT	CTGTTTTTCAT	GATTTTCCTC	TGCGCGTATC
CTGGGCCGTA	GCGTCGGCAT	CAGCCGACAA	GACAAAAGTA	CTAAAAGGAG	ACGCGCATAG
250	260	270	280	290	300
TCATCCGTTA	CCGGGAATTC	TTCAAAGACT	CCGTAATCGA	CCTCCTTACC	TGCCGATGGG
AGTAGGCAAT	GGCCCTTAAG	AAGTTTCTGA	GGCATTAGCT	GGAGGAATGG	ACGGCTACCC
310	320	330	340	350	360
TTGCGTACTG	CAGCTGCAGC	TGTAAGTGCA	GCTGCAAATG	CATCTCGGGC	CCCTGTAGCC
AAGCGATGAC	GTCGACGTCG	ACATTACAGT	CGACGTTTAC	GTAGAGCCCG	GGGACATCGG
370	380	390	400	410	420
GCTGCTGTTT	AGCGTGTTAC	AAGGAGACGA	TGATTTACGA	CATGGTCCAA	TACGGTCATC
CGACGACAAG	TCCGACAATG	TTCTCTGCT	ACTAAATGCT	GTACCAGGTT	ATGCCAGTAG
430	440	450	460	470	480
GACGGCGTCC	CGGACACGGC	GACGATCCCG	ACAGGGTGAT	CTGCGAGATA	GTCGAGAGTC
CTGCGCGAGG	GCCTGTGCCG	CTGCTAGGGC	TGTCCCACTA	GACGCTCTAT	CAGCTCTCAG
490	500	510	520	530	540
CCCCGGTTTC	GGCGCCGACG	GTGTCCGTCC	CCCCGCCGTC	GGAGGAGTCC	CACCAGCCCG
GGGGCCAAAG	CCGCGGCTGC	CACAGGCAGG	GGGGCGGCAG	CCTCCTCAGG	GTGGTCGGGC
550	560	570	580	590	600
TCATCCCACC	GCAGCCGCCA	GCACCGACAT	CGGAACCCAA	ACCGAAGAAA	GGTAGGGCGA
AGTAGGGTGG	CGTCGGCGGT	CGTGGCTGTA	GCCTTGGGTT	TGGCTTCTTT	CCATCCCGCT
610	620	630	640	650	660
AAGATAAACC	GAAGGGTAGA	CCGAAAGACA	AACCTCCGTG	CGAACCGACG	GTGAGTTTAC
TTCTATTGTT	CTTCCCATCT	GGCTTCTGTT	TTGGAGGCAC	GCTTGGCTGC	CACTCAAGTG
670	680	690	700	710	720
AACCACCGTC	GCAGCCGACG	GCAATGCCCC	GCGGTCCGCC	CGACGCGCCT	CCCCCCGCCA
TTGGTGGCAG	CGTCGGCTGC	CGTTACGGGC	CGCCAGGCGG	GCTGCGCGGA	GGGGGGCGGT
730	740	750	760	770	780
TGCCCGCAGAT	GCCACCCGGC	GTGGCCGAGG	CGGTACAAGC	TGCCGTGCAG	GCGGCCGTGG
ACGGCGTCTA	CGGTGGGCCG	CACCGGCTCC	GCCATGTTTC	ACGGCACGTC	CGCCGGCACC
790	800	810	820	830	840
CCGCGGCTCT	ACAACAACAG	CAGCAGCATC	AGACCGGAAC	GTAACCCGCC	CCCGGTGCGA
GGCGCCGAGA	TGTTGTTGTC	GTCGTCGTAG	TCTGGCCTTG	CATTGGGCGG	GGGCCACGCT
850	860	870	880	890	900
TAAGGAATTT	TCCGACTTGG	CGCACATCTC	CTTCTCTCAAT	GTTTGGACAA	TAAACACATT
ATTCTTTAAA	AGGCTGAACC	GCGTGTAGAG	GAAGGAGTTA	CAAACCTGTT	ATTGTGTAA
910	920	930	940	950	960
CCTTGCCAAA	AAATGACGTT	TCCAGAAATC	CAAGGCATAA	ATGTCCGTAC	ACCGGCCCTT
GGAACGGTTT	TTTACTGCAA	AGGTCTTTAG	GTTCCGTATT	TACAGGCATG	TGGCCGGGAA
970	980	990	1000	1010	1020
CCCAACACGG	AGTTTGTAGAT	TCCAAGCAGG	AGAGAAGATC	ATGGTGTGGA	TATGGCTCGG
CGGTCGCTGC	GCAAAAGTGA	ACGTTGCTGC	TCGTCCTTAC	TACGACACCT	ATACCGACCC

Fig. 1A

SEQ No 1
18318 base

1030	1040	1050	1060	1070	1080
CATCGGGCTC	CTCGGCGGTA	CCGGACTGGC	TTCCCTGGTC	CTGGCCATTT	CCTTATTTAC
GTAGCCCGAG	GAGCCGCCAT	GGCCTGACCG	AAGGGACCAG	GACCGGTAAA	GGAATAAATG
1090	1100	1110	1120	1130	1140
CCAGCGCCGA	GGCCGCAAGC	GATCCGACGA	GACTTCGTCT	CGAGGCCGGC	TCCCGGGTGC
GGTCGCGGCT	CCGGCGTTCT	CTAGGCTGCT	CTGAAGCAGC	GCTCCGGCCG	AGGGCCACG
1150	1160	1170	1180	1190	1200
TGCTTCTGAT	AAGCGTGGTG	CCTGCGCGTG	CTGCTATCGA	AATCCGAAAG	AAGACGTCTG
ACGAAGACTA	TTCGCACCAC	GGACGCGCAC	GACGATAGCT	TTAGGCTTTC	TTCTGCAGCA
1210	1220	1230	1240	1250	1260
CGAGCCGCTG	GATCTGGAAC	TGGGGCTCAT	GCGGGTGGAC	ACCCACCCGC	CGACGCCGCA
GCTCGGCGAC	CTAGACCTTG	ACCCCGAGTA	CGCCACCTG	TGGGTGGGCG	GCTCGGCGCT
1270	1280	1290	1300	1310	1320
GGTGCCGCGG	TGTACGTGCG	TCTACATAGG	AGAGGATGGT	CTGCCGATAG	ATAAACCCGA
CCACGGCGCC	ACATGCAGCG	AGATGTATCC	TCTCTACCA	GACGGCTATC	TATTTGGGCT
1330	1340	1350	1360	1370	1380
GTTCCTCCG	GCGCGGTTCT	AGATCCCGGA	CGTATCCACG	CCGGGAACGC	CGACCAGCAT
CAAAGGAGGC	CGCGCCAAGC	TCTAGGGGCT	GCATAGGTGC	GGCCCTTGCG	GCTGGTCTGA
1390	1400	1410	1420	1430	1440
CGGCCGATCT	CCGTCTGATT	GCTCCTCGTC	GAGCTCTTTG	TCGTCTCTGA	CCAGCGTCTGA
GCCGGCTAGA	GGCAGCGTAA	CGAGGAGCAG	CTCGAGAAAC	AGCAGGAGCT	GGTCGCAGCT
1450	1460	1470	1480	1490	1500
CACGGTGCTG	TATCAGCCGC	CGCCATCCTG	GAAGCCACCT	CCGCCGCCCG	GGCGCAAGAA
GTGCCACGAC	ATAGTCGGCG	GCGGTAGGAC	CTTCGGTGGA	GGCGGCGGGC	CCGCGTTCTT
1510	1520	1530	1540	1550	1560
GCGGCCGCTT	ACGCCGCGCG	TCCGGGCCCC	CACCACGCGG	CTGTCTGTCT	ACAGACCCCC
CGCCGGCGGA	TGCGGCGGCC	AGGCCCGGGG	GTGGTGCGCC	GACAGCAGCG	TGTCTGGGGG
1570	1580	1590	1600	1610	1620
GACGCCGATA	CCCGCGCCGC	GTAAGAACCT	GAGCACGCGC	CCCACCAAGA	AAACGCCGCC
CTGCGGCTAT	GGGCGCGGCG	CATTCTTGGA	CTCGTGCGGC	GGGTGGTTCT	TTTGCGGCGG
1630	1640	1650	1660	1670	1680
GCCCACGAAA	CCCAAGCCGG	TCCGCTGGAC	ACCGCCGGTG	ACACCCAGGC	CCTTCCCAGAA
CGGGTGCTTT	GGGTTCGGCC	AGCCGACCTG	TGGCGGCCAC	TGTGGGTCCG	GGAAGGGCTT
1690	1700	1710	1720	1730	1740
AACGCCGACG	CCACAAAAGC	CGCCGCGGAA	TCCGAGACTA	CCGCGCACCG	TCCGTCTGGA
TTGCGGCTGC	GGTGTTTTCT	GCGGCGCCTT	AGGCTCTGAT	GGCGCGTGCC	AGCCAGACCT
1750	1760	1770	1780	1790	1800
GAATCTCTCG	AAGGTGGGAC	TCTCGTGTCC	CTGTCCCGCA	CCCCGCACGC	CGACGGAGCC
CTTAGAGAGC	TTCCACCCTG	AGAGCACAGG	GACAGGGGCT	GGGGCGTGCG	GCTGCCTCGG
1810	1820	1830	1840	1850	1860
GACCACGCTG	CCTATCGTGT	CGGTTTCCGA	GCTAGCCCCG	CCTCCTCGAT	GGTCGGACAT
CTGGTGCGAC	GGATAGCACA	GCCAAAGGCT	CGATCGGGGC	GGAGGAGCTA	CCAGCCTGTA
1870	1880	1890	1900	1910	1920
CGAGGAACTC	TTGGAACAGG	CGGTGCAGAG	CGTCATGAAG	GACGCCGAGT	CGATGCAGAT
GCTCCTTGAG	AACCTTGTTC	GCCACGTCTC	GCAGTACTTC	CTGCGGCTCA	GCTACGTCTA
1930	1940	1950	1960	1970	1980
GACCTGAGAC	CGAAAGAGCG	AGCGCGTCCG	TTGTACAGTT	GTATAGCAGC	ACACGCCTTC
CTGGACTCTG	GCTTTCTCTG	TGCGCGCAGC	AACATGTCAA	CATATCGTCT	TGTGCGGAAG
1990	2000	2010	2020	2030	2040
CCTCTTTTTC	ACCGCAGCTA	AGAGAGAGAA	AGAGAGTATG	TCAGTCAAGG	GCGTGGAGAT
GGAGAAAAAG	TGGCGTCGAT	TCTCTCTCTT	TCTCTCATAC	AGTCAGTTCC	CGCACCTCTA

Fig. 1B

1000	2050	2060	2070	2080	2090	2100
	GCCAGAAATG	ACGTGGGACT	TGGACGTTAG	AAATAAATGG	CGGCGTCGAA	AGGCCCTGAG
	CGGTCITTTAC	TGCACCCCTGA	ACCTGCAATC	TTTATTTTACC	GCCGCAGCTT	TCCGGGACTC
	2110	2120	2130	2140	2150	2160
	TCGCATTAC	CGGTTCTGGG	AATGTCGGCT	ACGGGTGTGG	TGGCTGAGTG	ACGCCGGCGT
	AGCGTAAGTG	GCCAAGACCC	TTACAGCCGA	TGCCCACACC	ACCGACTCAC	TGCGGCCGCA
	2170	2180	2190	2200	2210	2220
	AAGAGAAACC	GACCCACCCG	GTCCCCGACG	CCGCCCGACT	TGGATGACCG	CGGTGTTTCA
	TTCTCTTTGG	CTGGGTGGCG	CAGGGGCTGC	GGCGGGCTGA	ACCTACTGTC	GCCACAAAGT
	2230	2240	2250	2260	2270	2280
CGTTATCTGT	GCCGTTTTTC	TTACGCTTAT	GATTATGGCC	ATCGGCGCGC	TCATCGCGTA	
GCAATAGACA	CGGCAAAACG	AATGCGAATA	CTAATACCGG	TAGCCGCGCG	AGTAGCGCAT	
2000	2290	2300	2310	2320	2330	2340
	CTTAAGATAT	TACCACCAGG	ACAGTTGGCG	AGACATGCTC	CACGATCTAT	TTTGCGGCTG
	GAATTCTATA	ATGGTGGTCC	TGTCAACCCG	TCTGTACGAG	GTGCTAGATA	AAACGCCGAC
	2350	2360	2370	2380	2390	2400
	TCATTATCCC	GAGAAGTGCC	GTCCGGCACCA	CGAGCGGCAG	AGAAGGAGAC	GGCAAGCCAT
	AGTAATAGGG	CTCTTCACGG	CAGCCGTGGT	GCTCGCCGTC	TCTTCCTCTG	CCGTTCGGTA
	2410	2420	2430	2440	2450	2460
	GGATGTGCCC	GACCCGGAAC	TCGGCGACCC	GGCCCCGCCG	CCGTTGAACG	GAGCTATGTA
	CCTACACGGG	CTGGGCCTTG	AGCCGCTGGG	CCGGCGCGCC	GGCAACTTGC	CTCGATACAT
	2470	2480	2490	2500	2510	2520
CTACGGCAGC	GGCTGTCGCT	TCGACACGGT	GGAATTGGTG	GACGAGACGA	GACCCGCGCC	
GATGCCGTCG	CCGACAGCGA	AGCTGTGCCA	CCTTTACCAC	CTGCTCTGCT	CTGGGCGCGG	
3000	2530	2540	2550	2560	2570	2580
	GCCGCGCGCTG	TCATCGCCCG	AAACCGGCGA	CGATAGCAAC	GACGACGCGG	TTGCGGCGGG
	CGGCGCGCGAC	AGTAGCGGGC	TTTGGCCGCT	GCTATCGTTG	CTGCTGCGCC	AACGCGCGCC
	2590	2600	2610	2620	2630	2640
	AGGTGCTGGC	GGGGTAACAT	CACCCGCGAC	TCGTACGACG	TCGCCGAACG	CACTGCTGCC
	TCCACGACCG	CCCCATTGTA	GTGGGCGCTG	AGCATGCTGC	AGCGGCTTGC	GTGACGACGG
	2650	2660	2670	2680	2690	2700
	AGAATGGATG	GATGCGGTGC	ATGTGGCGGT	CCAAGCCGCC	GTTCAAGCGA	CCGTGCAAGT
	TCTTACCTAC	CTACGCCACG	TACACCGCCA	GGTTGCGCGG	CAAGTTGCGT	GGCACGTTCA
	2710	2720	2730	2740	2750	2760
AAGTGGCCCG	CGGGAGAACG	CCGTATCTCC	CGCTACGTAA	GAGGGTTGAG	GGGGCCGTTT	
TTACACGGGC	GCCCTCTTTG	GGCATAGAGG	GCGATGCATT	CTCCCAACTC	CCCCGGCAAG	
4000	2770	2780	2790	2800	2810	2820
	CCGCGCGAGT	GCTGTACAAA	AGAGAGAGAC	TGGGACGTAG	ATCCGGACAG	AGGACGGTCA
	GGCGCGCTCA	CGACATGTTT	TCTCTCTCTG	ACCCTGCATC	TAGGCTGTTC	TCCTGCCAGT
	2830	2840	2850	2860	2870	2880
	CCATGGACGA	TCTGCCGCTG	AATGTCGGGT	TACCCATCAT	CGGCGTGATG	CTCGTGCTGA
	GGTACCTGCT	AGACGCGGAC	TTACAGCCCA	ATGGGTAGTA	GCCGCACTAC	GAGCACGACT
	2890	2900	2910	2920	2930	2940
	TCGTGGCCAT	CCTCTGCTAT	CTGGCTTACC	ACTGGCACGA	CACCTTCAAA	CTGGTGCGCA
	AGCACCGGTA	GGAGACGATA	GACCGAATGG	TGACCGTGCT	GTGGAAGTTT	GACCACGCGT
	2950	2960	2970	2980	2990	3000
TGTTTTCTGAG	CTACCCTGTG	CTGATCCGCT	GTTGCGAGCT	GTACGGGGAG	TACGAGCGCC	
ACAAAGACTC	GATGGCGACC	GACTAGGCGA	CAACGCTCGA	CATGCCCTTC	ATGCTCGCGG	
5000	3010	3020	3030	3040	3050	3060
	GGTTCGCGGA	CCTGTGCTCT	CTGGGCCTCG	GCGCCGTACG	GCGGAGTCG	GACAGACGAT
	CCAAGCGCCT	GGACAGCAGA	GACCCGGGAG	CGCGGCATGC	CGCCCTCAGC	CTGTCTGCTA

Fig. 1C

3070	3080	3090	3100	3110	3120
ACCGTTTCTC	CGAACGGCCC	GACGAGATCT	TGGTCCGTTG	GGAGGAAGTG	TCTTCCCAGT
TGGCAAAGAG	GCTTGCCGGG	CTGCTCTAGA	ACCAGGCAAC	CCTCCTTCAC	AGAAGGGTCA
3130	3140	3150	3160	3170	3180
GCAGCTACGC	GTCGTGCGGG	ATAACAGACC	GCCGTGTGGG	TTCATCGTCT	TCGTGCTCGG
CGTCGATGCG	CAGCAGCGCC	TATTGTCTGG	CGGCACACCC	AAGTAGCAGA	AGCAGCAGCC
3190	3200	3210	3220	3230	3240
TCCACGTGCG	TAGCCAGAGA	AACAGCGTGC	CTCCGCCGGA	CATGGCGGTG	ACGGCGCCGC
AGGTGCAGCG	ATCGGTCTCT	TTGTGCGACG	GAGGCGGCCT	GTACCGCCAC	TGCCGCGGCG
3250	3260	3270	3280	3290	3300
TGACCGACGT	CGATCTGTTG	AAACCCGTGA	CGGGATCCGC	GACGCAGTTC	ACCACCGTAG
ACTGGCTGCA	GCTAGACAAC	TTTGGGCACT	GCCCTAGGCG	CTGCGTCAAG	TGGTGGCATC
3310	3320	3330	3340	3350	3360
CCATGGTACA	TTATCATCAA	GAGTACACGT	GAATGAGAAA	AAGAAAAAAG	AGGGGAGCGG
GGTACCATGT	AATAGTAGTT	CTCATGTGCA	CTTACTCTTT	TTCTTTTTTC	TCCCCTCGCC
3370	3380	3390	3400	3410	3420
ATCGCGATAA	TGTCGCTTTG	ACATTCTCTG	CTCGATCTAC	TCAGCGTCTG	CACGAAACGG
TAGCGCTATT	ACAGCGAAAC	TGTAAGAGAC	GAGCTAGATG	AGTCGCAGAC	GTGCTTTGCC
3430	3440	3450	3460	3470	3480
CATCCGCACG	GAGGCGAGCC	CAAGCGTATC	TGCAGCAAGC	GGTTCTTTCC	CTCGGTGATG
GTAGGCGTGC	CTCCGCTCGG	GTTGCGCATAG	ACGTGCTTCG	CCAAGAAAGG	GAGCCACTAC
3490	3500	3510	3520	3530	3540
GTGGCAGCAT	CGGTGCGGGG	AGCTTGTTTCG	GACGATGGAC	GGTGAGGAGT	CCCTGGCGAT
CACCGTCGTA	GCCACCGCCC	TGGAACAAGC	CTGCTACCTG	CCACTCCTCA	GGGACCGCTA
3550	3560	3570	3580	3590	3600
CAGGCGGCTC	CCGGGTGTGG	AGTTCAACGG	GTGGTAATGG	TGGCGGTGAT	CGGTGTTAGA
GTCCGCCGAG	GGCCACACCC	TCAAGTTGCC	CACCATTACC	ACCGCCACTA	GCCACAATCT
3610	3620	3630	3640	3650	3660
AAACGTTGGC	CCTGGCAAAC	ATATATCTAC	TGTAAACCCCT	CTGCTCTGTT	AATAAAAAGC
TTTGCCACCG	GGACCGTTTG	TATATAGATG	ACATTTGGGA	GACGAGACAA	TTATTTTTTCG
3670	3680	3690	3700	3710	3720
ACACTTTTCA	CATGAGTTTCG	TAATTTTATT	GTGTAGTGGG	AATTTTTTACG	TCATTGGGAA
TGTGAAAAGT	GTACTCAAGC	ATTAAAAATA	CACATCACCT	TTAAAAATGC	AGTAACCCTT
3730	3740	3750	3760	3770	3780
ACCCCAAGAT	GAAAGAGTAT	AATGTGCATA	TCACCGGGGG	TTCCCTGTCA	GTACGAATGT
TGGGGTCTTA	CTTTCTCATA	TTACACGTAT	AGTGGCCCCC	AAGGGACAGT	CATGCTTACA
3790	3800	3810	3820	3830	3840
ACACAACGCG	GGTTACATTA	CGATAAACTT	TCCGGTAAAA	CGATGCCGAT	ACAGCGTGTA
TGTGTTGCGC	CCAATGTAAT	GCTATTTGAA	AGGCCATTTT	GCTACGGCTA	TGTGCGACAT
3850	3860	3870	3880	3890	3900
TAACGCTGAT	TGTTACGACA	AACGAGTTGG	TATATCCATT	ATATAGTAAC	GAACATGCTG
ATTGCGACTA	ACAATGCTGT	TTGCTCAACC	ATATAGGTAA	TATATCATTG	CTTGTACGAC
3910	3920	3930	3940	3950	3960
TGGATATTAG	TTTTATTTCG	ACTCGCCGCA	TCGGCGAGTG	AAACCACTAC	AGGTACCAGC
ACCTATAATC	AAAATAAACG	TGAGCGGCGT	AGCCGCTCAC	TTTGGTGATG	TCCATGGTCC
3970	3980	3990	4000	4010	4020
TCTAATTCCA	GTCAATCTAC	TAGTGCTACC	GCCAACACGA	CCGTATCGAC	ATGTATTAAAT
AGATTAAAGT	CAGTTAGATG	ATCAGCATGG	CGGTTGTGCT	GGCATTAGCTG	TACATAATTA
4030	4040	4050	4060	4070	4080
GCCTCTAACG	GCAGTAGCTG	GACAGTACCA	CAGCTGCGCG	TGCTTGCCGC	TAGCGGCTGG
CGGAGATTGC	CGTCATCGAC	CTGTCATGTT	GTCGAGCGCG	ACGAACGGCG	ATCGCCGACC

Fig. 1D

4090 4100 4110 4120 4130 4140
ACATTATCTG GACTCCTTCT CTTATTTTACC TGCTGCTTTT GCTGCTTTTG GCTAGTACGT
TGTAATAGAC CTGAGGAAGA GAATAAATGG ACGACGAAAA CGACGAAAAC CGATCATGCA

4150 4160 4170 4180 4190 4200
AAAATCTGCA GCTGCTGCGG CAACTCCTCC GAGTCAGAGA GCAAAACAAC CCACGCGTAC
TTTTAGACGT CGACGACGCC GTTGAGGAGG CTCAGTCTCT CGTTTTGTTG GGTGCGCATG

4210 4220 4230 4240 4250 4260
ACCAATGCCG CATTCACCTT TTCCGACGCA ACGTTACCCA TGGGCACTAC AGGGTCGTAC
TGTTTACGGC GTAAGTGAAG AAGGCTGCGT TGCAATGGGT ACCCGTGATG TCCCAGCATG

4270 4280 4290 4300 4310 4320
ACTCCCCAC AGGACGGCTC ATTTCACCTT CCGCCTCGGT GACGTAGGCT AAACCGAAAC
TGAGGGGGTG TCCTGCCGAG TAAAGGTGGA GGCGGAGCCA CTGCATCCGA TTTGGCTTTG

4330 4340 4350 4360 4370 4380
CCACGTTGAA CCTAACGCGG TTTCGGAAGG CCTGAGACGT CACTTTCACA ATGACGTCCG
GGTGCAACTT GGATTGCGCC AAAGCCTTCC GGACTCTGCA GTGAAAGTGT TACTGCAGGC

4390 4400 4410 4420 4430 4440
TATACACGTT CATCATAAAA CACCGTAGAG GCTAAGGCTT CCGTAGGGAG AGACCTCAAC
ATATGTGCAA GTAGTATTTT GTGGCATCTC CGATTCCGAA GCCATCCCTC TCTGGAGTTG

4450 4460 4470 4480 4490 4500
TGTTCTCGAT GAGCACCCGT GCTCTCATCT CTTCAGACTT GTCATGACCC CCGCTCAGAC
ACAAGGACTA CTCGTGGGCA CGAGAGTAGA GAAGTCTGAA CAGTACTGGG GGCGAGTCTG

4510 4520 4530 4540 4550 4560
TAACGCGACT ACCACCGTGC ACCCGCACGA CGCAAAAAAC GGCAGCGGCG GTAGTGCCCT
ATTGCGCTGA TGTGCGCAGG TGGGCGTGCT GCGTTTTTTG CCGTCGCCGC CATCACGGGA

4570 4580 4590 4600 4610 4620
GCCGACCCTC GTCGTTTTCG GCTTTATCGT TACGCTACTT TTCTTTCTCT TTATGCTCTA
CCGGCTGGGAG CAGCAAAAGC CGAAATAGCA ATGCGATGAA AAGAAAGAGA AATACGAGAT

4630 4640 4650 4660 4670 4680
CTTTTGAAC AACGACGTGT TCCGTAAGCT GCTCCGTGCG CTTGGATCCA GCGCTGTTGC
GAAAACCTTG TTGCTGCACA AGGCATTGCA CGAGGCACGC GAACCTAGGT CGCGACAACG

4690 4700 4710 4720 4730 4740
GACCGCTTCG ACGCGTGGCA AGACGAGGTC ATCTACCGTC GTCCATCAGC TCGTTCCCAG
CTGGCGAAGC TGCGCACCGT TCTGCTCCAG TAGATGGCAG CAGGTAGTGC AGCAAGGGTC

4750 4760 4770 4780 4790 4800
AGCGACGACG AGAGTCGTAC TAACAGCGTG TCATCGTACG TTCTTTTATC ACCCGCGTCC
TCGCTGCTGC TCTCAGCATG ATTGTGCGAC AGTAGCATGC AAGAAAATAG TGGGCGCAGG

4810 4820 4830 4840 4850 4860
GATGGCGGTT TTGACAACCC GGCAGTGACA GAGGCCGTCG ACAGCGTGA GACTGGGCG
CTACCGCAA AACTGTTGGG CCGTGACTGT CTCCGGCAGC TGTCGCACCT GCTGACCCGC

4870 4880 4890 4900 4910 4920
ACCACCTCGG TTTTCTACGC CACGTCCGAC GAAACGGCGG ACGCCGAGCG CCGAGACTCG
TGGTGGAGCC AAAAGATGCG GTGCAGGCTG CTTTGCCGCC TGCGGCTCGC GGCTCTGAGC

4930 4940 4950 4960 4970 4980
CAGCAACTGC TCATCGAGCT TCCGCCGGAG CCGCTCCCGC CCGACGTGGT GGCGGCCATG
GTGCTTGACG AGTAGCTCGA AGGCGGCCCTC GGCGAGGGCG GGCTGCACCA CCGCCGGTAC

4990 5000 5010 5020 5030 5040
CAGAAAGCAG TGAAACGCGC TGTACAGAAC GCACTACGAC ACAGCCACGA CTCTTGGCAG
GTCTTTCGTC ACTTTCGCGC ACATGTCTTG CGTGATGCTG TGTCGGTGCT GAGAACCCTC

5050 5060 5070 5080 5090 5100
CTTCATCAGA CCCTGTGACG CCAGATGAAC GTTCCTTCTT AAACATCCGA GGTAGCAATG
GAAGTAGTCT GGGACACTGC GGTCTACTTG CAAGGAAGAA TTTGTAGGCT CCATCGTTAC

Fig. 1E

5110 5120 5130 5140 5150 5160
AGACAGGTCG CGTACCGCCG GCGACGCGAG AGTTCTCTGCG CGGTGCTGGT CCACCACGTC
TCTGTCCAGC GCATGGCGGC CGCTGCGCTC TCAAGGACGC GCCACGACCA GGTGGTGCAG

5170 5180 5190 5200 5210 5220
GGCCGCGACG GCGACGCGCA GGGGGAGGCA GCAAAAAAGA CCTGCAAAAA AACCGGACGC
CCGGCGCTGC CGCTGCCGCT CCCCTCCGT CGTTTTTCTT GGACGTTTTT TTGGCCTGCG

5230 5240 5250 5260 5270 5280
TCAGTTGCGG GCATCCCGGG CGAGAAGCTG CGTCGCACGG TGGTCACCAC CACGCCGGCC
AGTCAACGCC CGTAGGGCCC GCTCTTCGAC GCAGCGTGCC ACCAGTGGTG GTGCGGCCGG

5290 5300 5310 5320 5330 5340
CGACGTTTGA GCGGCCGACA CACGGAGCAG GAGCAGGCGG GCATGCGTCT CTGTGAAAAA
GCTGCAAACT CGCCGGCTGT GTGCCTCGTC CTCGTCCGCC CGTACGCAGA GACACTTTTT

5350 5360 5370 5380 5390 5400
GGGAAGAAAA GAATCATCAT GTGCCGCCGG GAGTCGCTCC GAACTCTGCC GTGGCTGTTC
CCCTTCTTTT CTTAGTAGTA CACGGCGGCC CTCAGCGAGG CTTGAGACGG CACCGACAAG

5410 5420 5430 5440 5450 5460
TGGGTGCTGT TGAGCTGCCC GCGACTCCTC GAATATTCTT CCTCTTCGTT CCCCTTCGCC
ACCCACGACA ACTCGACGGG CGCTGAGGAG CTTATAAGAA GGAGAAGCAA GGGGAAGCGG

5470 5480 5490 5500 5510 5520
ACCGCTGACA TTGCCGAAAA GATGTGGGCC GAGAATTATG AGACCACGTC GCCGGCGCCG
TGCGGACTGT AACGGCTTTT CTACACCCGG CTCTTAATAC TCTGGTGCAG CGGCCGCGGC

5530 5540 5550 5560 5570 5580
GTGTTGGTGC CCGAGGGAGA GCAAGTTACC ATCCCCTGCA CGGTCAAGAC AACTCCTGG
CACAAACCAGC GGCTCCCTCT CGTTCAATGG TAGGGGACGT GCCAGTACTG TGTGAGGACC

5590 5600 5610 5620 5630 5640
CCCATGGTCT CCATTCGCGC ACGTTTCTGT CGTTCCACAG ACGGCAGCGA CGAGCTCATC
GGGTACCAGA GGTAAGCGCG TGCAAAGACA GCAAGGGTGC TGCCGTCGCT GCTCGAGTAG

5650 5660 5670 5680 5690 5700
CTGGACGCCG TCAAAGGCCA TCGGCTGATG AACGGACTCC AGTACCGCCT GCCGTACGCC
GACCTGCGGC AGTTTCCGGT AGCCGACTAC TTGCCTGAGG TCATGGCGGA CGGCATGCGG

5710 5720 5730 5740 5750 5760
ACTTGGGAATT TCTCGCAATT GCATCTCGGC CAAATAITCT CGCTTACTTT TAACGTATCG
TGAACCTTAA AGAGCGTTAA CGTAGAGCCG GTTTATAAGA GCGAATGAAA ATTGCATAGC

5770 5780 5790 5800 5810 5820
ATGGACACAG CCGGCATGTA CGAATGCGTG CTACGCAACT ACAGCCACGG CCTCATCATG
TACCTGTGTC GGGCGTACAT GCTTACGCAC GATGCGTTGA TGTCGGTGCC GGAGTAGTAC

5830 5840 5850 5860 5870 5880
CAACGCTTCG TAATTCTCAC GCAGCTGGAG ACGCTCAGCC GGCCCGACGA ACCTTGCTGC
GTTGCGAAGC ATTAAGAGTG CGTCGACCTC TGCGAGTCGG CCGGGCTGCT TGGAACGACG

5890 5900 5910 5920 5930 5940
ACACCGGCGT TAGGTCGCTA CTCGCTGGGA GACCAGATCT GGTGCGCGAC GCCCTGGCGT
TGTGCGCCGA ATCCAGCGAT GAGCGACCCT CTGGTCTAGA CCAGCGCTG CCGGACCGCA

5950 5960 5970 5980 5990 6000
CTACGGAATC ACGACTGCGG AACGTACCGC GGCTTTCAAC GCAACTACTT CTATATCGGC
GATGCCTTAG TGCTGACGCC TTGCATGGCG CCGAAAGTTG CGTTGATGAA GATATAGCCG

6010 6020 6030 6040 6050 6060
CGCGCCGACG CCGAGGATTG CTGGAACCC GCATGTCCGG ACGAGGAACC CGACCGCTGT
GCGCGGCTGC GGCTCCTAAC GACCTTTGGG CGTACAGGCC TGCTCCTTGG GCTGGCGACA

6070 6080 6090 6100 6110 6120
TGGACAGTGA TACAGCGTTA CCGGCTCCCC GCGGACTGCT ACCGTTGCA GCCACACCCG
ACCTGTCACT ATGTCGCAAT GGCCGAGGGG CCGCTGACGA TGGAAGCGT CGGTGTGGGC

Fig. 1F

6130	6140	6150	6160	6170	6180
CCGAAATTTT	TACCGGTGAC	GCCAGCACCG	CCGGCCGACA	TAGACACCGG	GATGTCTCCC
GGCTTTAAAA	ATGGCCACTG	CGGTCGTGGC	GGCCGGCTGT	ATCTGTGGCC	CTACAGAGGG
6190	6200	6210	6220	6230	6240
TGGGCCACTC	GGGGAATCGC	GGCGTTTTTG	GGGTTTTGGA	GTATTTTTTAC	CGTATGTTTC
ACCCGGTGAG	CCCCTTAGCG	CCGCAAAAAC	CCCAAAACCT	CATAAAAATG	GCATACAAAG
6250	6260	6270	6280	6290	6300
CTATGCTACC	TGTGTTATCT	GCAGTGTGT	GGACGCTGGT	GTCCCAACGC	GGGAAGGGGA
GATACGATGC	ACACAATAGA	CGTCACAACA	CCTGCGACCA	CAGGGTGCGG	CCCTTCCCCT
6310	6320	6330	6340	6350	6360
CGACGAGGCG	GTGAGGGCTA	TCGACGCCTA	CCGACTTACG	ATAGTTACCC	CGGTGTTAGA
GCTGCTCCGC	CACTCCCAGT	AGCTGCGGAT	GGCTGAATGC	TATCAATGGG	GCCACAATCT
6370	6380	6390	6400	6410	6420
AAGATGAAGA	GGTGAGAACA	CGTATAAAAT	AAAAAAATAA	TATGTTAAAA	AATGCAGTGT
TTCTACTTCT	CCACTCTTGT	GCATATTTTA	TTTTTTTATT	ATACAATTTT	TTACGTACAA
6430	6440	6450	6460	6470	6480
GTGAAGTGTG	AATAGTGTGA	TTAAAAATATG	CGGATTGAAT	GGGTGTGGTG	GTTATTCGGA
CACCTTCACAC	TTATCACACT	AATTTTATAC	GCCTAACTTA	CCCACACCAC	CAATAAGCCT
6490	6500	6510	6520	6530	6540
TACTTTGTGT	CATCCGTTGG	GAGCGAACGG	TCATTATCCT	ATCGTTACCA	CTTGGGAATCT
ATGAAACACA	GTAGGCAACC	CTCGCTTGCC	AGTAATAGGA	TAGCAATGGT	GAACCTTAGA
6550	6560	6570	6580	6590	6600
AATTCATCTA	CCAACGTGGT	TTGCAACGGA	AACATTTCCG	TGTTTGTAAA	CGGCACCCTA
TTAAGTAGAT	GGTTGCACCA	AACGTTGCCT	TTGTAAAGGC	ACAAACATTT	GCCGTGGGAT
6610	6620	6630	6640	6650	6660
GGTGTGCGGT	ATAACATTAC	GGTAGGAATC	AGTTCGTCTT	TATTAATAGG	ACACCTTACT
CCACACGCCA	TATTGTAAATG	CCATCCTTAG	TCAAGCAGAA	ATAATTATCC	TGTGGAATGA
6670	6680	6690	6700	6710	6720
TATCAAGTAT	TGGAATCATG	GTTACACCCC	TGGGTCCAAA	ATAAAAGTTA	CAACAAACAA
TATGTTTCATA	ACCTTAGTAC	CAAGTGTGGG	ACCCAGGTTT	TATTTTCAAT	GTTGTTTGTT
6730	6740	6750	6760	6770	6780
CCCCTAGGTG	ACACTGAAAC	GCTTTATAAT	ATAGATAGCG	AAAACATTCA	TCGCGTATCT
GGGGATCCAC	TGTGACTTTG	CGAAATATTA	TATCTATCGC	TTTTGTAAAGT	AGCGCATAGA
6790	6800	6810	6820	6830	6840
CAATATTTTC	ACACAAGATG	GATAAAATCT	CTGCAAGAGA	ATCACACTTG	CGACCTCACA
GTTATAAAAG	TGTGTTCTAC	CTATTTTAGA	GACGTTCTCT	TAGTGTGAAC	GCTGGAGTGT
6850	6860	6870	6880	6890	6900
AACAGTACAC	CTACCTATAC	ATATCAAGTA	AACGTGAACA	ACACGAATTA	CCTAACACTA
TTGTCAATGTG	GATGGATATG	TATAGTTTAT	TTGCACTTGT	TGTGCTTAAT	GGATTGTGAT
6910	6920	6930	6940	6950	6960
ACATCTCCGG	GATGGCAAGA	CCGTCTAAAT	TACACCGTCA	TAAATAGTAC	ACACTTTAAC
TGTAGGAGCC	CTACCGTTCT	GGCAGATTTA	ATGTGGCAGT	ATTTATCATG	TGTGAAATTG
6970	6980	6990	7000	7010	7020
CTCACAGAAT	CGAACATAAC	CAGCATTTCA	AAATATCTCA	ACACTACCTG	CATAGAAAGA
GAGTGTCTTA	GCTTGTATTG	GTCGTAAGTT	TTTATAGAGT	TGTGATGGAC	GTATCTTTCT
7030	7040	7050	7060	7070	7080
CTCCGTAACT	ACACCTTGGA	GTCCGTATAC	ACCACAACTG	TGCCTCAAAA	CATAACAACA
GAGGCATTGA	TGTGGAACCT	CAGGCATATG	TGGTGTGAC	ACCGAGTTTT	GTATTGTTGT
7090	7100	7110	7120	7130	7140
TCTCAACACG	CAACAACCAC	TATGCACACA	ATACCTCCAA	ATACAATAAC	AATTCAAAAT
AGAGTTGTGC	GTTGTTGGTG	ATACGTGTGT	TATGGAGGTT	TATGTTATTG	TTAAGTTTTA

Fig. 1G

7150	7160	7170	7180	7190	7200
ACAACTCAAA TGTTGAGTTT	GCCATACTGT CGGTATGACA	ACAGACGCCG TGTCTGCGGC	TCTTTTAACG AGAAAATTGC	ACACACATAA TGTGTGTATT	CGTGACGAAA GCACTGCTTT
7210	7220	7230	7240	7250	7260
CACACGTTAA GTGTGCAATT	ACATAAGCTA TGTATTTCGAT	CGTTTTATCA GCAAAATAGT	CAAAAAACGA GTTTTTTGCT	ATAACACAAC TATTGTGTTG	ATCACCCTGG TAGTGGCACC
7270	7280	7290	7300	7310	7320
ATATATGCCA TATATACGGT	TACCTATGGG ATGGATACCC	CGCTACAGCC GCGATGTCGG	ACAATAGGCG TGTTATCCGC	CCGGTTTATA GGCCAAATAT	TATCGGGAAA ATAGCCCTTT
7330	7340	7350	7360	7370	7380
CACTTTACGC GTGAAATGCG	CGGTTAAGTT GCCAATTCAA	CGTATACGAG GCATATGCTC	GTATGGCGCG CATACCGCGC	GTCAGTAAAG CAGTCATTTT	ACGATTCGGA TGCTAAGCCT
7390	7400	7410	7420	7430	7440
TTCAACACAT AAGTTGTGTA	ATACTCCCCA TATGAGGGGT	CGATCCTCGA GCTAGGAGCT	ACACCTTACA TGTGGAATGT	GCATATGAGC CGTATACTCG	AAAAACAAG TTTTTTGTTT
7450	7460	7470	7480	7490	7500
AAAGTATAGC TTTCATATCG	CACAATCACA GTGTTAGTGT	TTTGGGCGAA AAACCCGCTT	TAACATGCTG ATTGTACGAC	TCATCCACTA AGTAGGTGAT	GCGTCTATTA CGCAGATAAT
7510	7520	7530	7540	7550	7560
ATCTAATGTT TAGATTACAA	TAACGGGAGC ATTGCCCTCG	TGTACTGTCA ACATGACAGT	CCGTTAAAAT GGCAATTTTA	ATCCATGGGA TAGGTACCCT	ATCAACGGGT TAGTTGCCCA
7570	7580	7590	7600	7610	7620
CAACCAACGT GTTGGTTGCA	CCATCAGCTT GGTAGTCGAA	GTGATTGTGC CACTAACACG	TCCATCTGGG AGGTAGACCC	TAACCGCTGT ATTGGCGACA	CAGCCTTGGC GTCGGAACCG
7630	7640	7650	7660	7670	7680
GACAGGTGTA CTGTCCACAT	ATCAGAGCTG TAGTGTGAC	TCACATAACT AGTGTATTGA	CACGAAGCCT GTGCTTCGGA	CCATCACAG GGTTAGTGTC	CAGCACACAT GTCGTGTGTA
7690	7700	7710	7720	7730	7740
AGTCCTAACG TCAGGATTGC	CCATTGGCGT GGTAACCGCA	GTATAAAAGT CATATTTTCA	TCGGAAAACT AGCCTTTTGA	TGACGGTTGT ACTGCCAACA	ACGGCACGAC TGCCGTGCTG
7750	7760	7770	7780	7790	7800
AAATCGATGT TTTAGCTACA	AGTGGTATGT TCACCATACA	TTTTCCAGCA AAAAGGTCGT	GAGACCGTGT CTCTGGCACA	GCGGTCTCTT CGCCAGAGAA	AGGTTTCGTA TCCAAGCGAT
7810	7820	7830	7840	7850	7860
TACTGTGGCT ATGACACCGA	GGAAACTGGT CCTTTGACCA	TACCTGTGAA ATGGACACTT	GATGGCTAAC CTACCGATTG	TATCCTGTTT ATAGGACAAG	TGTCCTGGAA ACAGGACCTT
7870	7880	7890	7900	7910	7920
AAACTTTTGG TTTGAAAACC	CGTCGTAGGT GCAGCATCCA	GGACTTTGCA CCTGAAACGT	GTATGCGGGT CATACGCCCA	TAGTGAAGTT ATCACTTCAA	ATGTCATTTA TACAGTAAAT
7930	7940	7950	7960	7970	7980
TTTACGTTTA AAATGCAAAT	CGATCTCGTA GCTAGAGCAT	TTACAAACCG AATGTTTGCC	CGGAGAGGAT GCCTCTCCTA	GATACCGTTC CTATGGCAAG	GGCCCCATGA CCGGGGTACT
7990	8000	8010	8020	8030	8040
GTTATTTTTA CAATAAAAAT	TTCTTCCGGT AAGAAGGCCA	AGGAGGCATG TCCTCCGTAC	AAGCCTCTGA TTCGGAGACT	TAATGCTCAT ATTACGAGTA	CTGCTTTGCT GACGAAACGA
8050	8060	8070	8080	8090	8100
GTGATATTAT CACTATAATA	TGCAGCTTGG ACGTCGAACC	AGTACTAAA TCACTGATTT	GTGTGTCAGC CACACAGTCG	ATAATGAAGT TATTACTTCA	GCAACTGGGC CGTTGACCCG
8110	8120	8130	8140	8150	8160
AATGAGTGCT TTACTCACGA	GCCCTCCGTG CGGGAGGCAC	TGGTTCGGGA ACCAAGCCCT	CAAAGAGTTA GTTTCTCAAT	CTAAAGTATG GATTTTCATAC	CACGGATTAT GTGCCTAATA

Fig. 1H

8170 ACCAGTGTAA TGGTCACATT	8180 CGTGTACCCC GCACATGGGG	8190 TTGCCCAAC AACGGGGTTG	8200 GGCACGTATG CCGTGCATAC	8210 TATCGGGACT ATAGCCCTGA	8220 TTACAACTGT AATGTTGACA
8230 ACCGATTGCA TGGCTAACGT	8240 CTCAATGTAA GAGTTACATT	8250 CGTCACTCAG GCAGTGAGTC	8260 GTCATGATTG CAGTACTAAG	8270 GTAAGTGCAC CATTGACGTG	8280 TTCCACCAAT AAGGTGGTTA
8290 AATACCGTAT TTATGGCATA	8300 GCGCACCTAA CGCGTGGAAT	8310 GAACCATACG CTTGGTATGC	8320 TACTTTTCCA ATGAAAAGGT	8330 CTCCAGGCGT GAGGTCCGCA	8340 CCAACATCAC GGTTGTAGTG
8350 AAACAACGAC TTTGTGCTG	8360 AGCAAAATCA TCGTTTTAGT	8370 TACCGCACAT ATGGCGTGTA	8380 ATAACCGTCA TATTGGCAGT	8390 AACAAGGAAA TTGTTCCFTT	8400 AAGCGGTCGT TTCGCCAGCA
8410 CATACTCTAG GTATGAGATC	8420 CCTGGTTGTC GGACCAACAG	8430 TCTCTTTATC AGAGAAATAG	8440 TTTCTTGTGG AAAGAACACC	8450 GTATCATACT CATAGTATGA	8460 TTTAATTCTC AAATTAAAGAG
8470 TATCTTATAG ATAGAATATC	8480 CCGCCTATCG GGCGGATAGC	8490 GAGTGAGAGA CTCACTCTCT	8500 TGCCAACAGT ACGGTTGTCA	8510 GTTGCTCAAT CAACGAGTTA	8520 CGGCAAAATT GCCGTTTTAA
8530 TTCTACCGCA AAGATGGCGT	8540 CCCTGTAAGC GGGACATTCG	8550 TTCCTGTGTG AAGGACAACA	8560 TGTPTTTTACA ACAAAAATGT	8570 TCACGGTACG AGTGCCATGC	8580 ATGAAGTCAC TACTTCAGTG
8590 ACAGATAATT TGCTCTATTAA	8600 ACAGATGAGC TGTCTACTCG	8610 TGTTTCATATT ACAAGTATAA	8620 TTTTATTATT AAAATAATAA	8630 TTTTCCAATT AAAAGGTTAA	8640 CCTGCACTAA GGACGTGATT
8650 AAAAAGAAGC TTTTTCTTCG	8660 ACTTTACGGA TGAAATGCCT	8670 ACCGTGTCTG TGGCACAGAC	8680 AGTATCTGTG TCATAGACAC	8690 GGGAATTTAG CCCTTAAATC	8700 GTACTTTTTG CATGAAAAAC
8710 CCGACGTCAG GGCTGCAGTC	8720 GAAAAATAAG CTTTTTTATC	8730 TGTGCGCTAC ACAGCGGATG	8740 ATAAGAGCCC TATTCTCGGG	8750 GGTGCTATCG CCACGATAGC	8760 TGCTGTCACT ACGACAGTGA
8770 CTTCTTGTGT GAAAGAACAA	8780 GCCTTCGATG CGGAAGCTAC	8790 TACGGCGTCC ATGCCGCAGG	8800 TGGCTCATTG ACCGAGTAAT	8810 CTACTCCTTC GATGAGGAAG	8820 ATCAGTAGCC TAGTCATCGG
8830 CCAGCGTTAT GGTCGCAATA	8840 GGTTAATTTT CCAATTAAAA	8850 AAGCATCATA TTCGTAGTAT	8860 ACGCCGTGCA TGCGGCACGT	8870 GCTGTTATGT CGACAATACA	8880 GCACGGACCC CGTGCCTGGG
8890 GAGACGCACT CTCTGCGTGA	8900 GCCGGATGGG CGGCCTACCC	8910 AACGTTTAAC TTGCAAAATTG	8920 CCATCATGCG GGTAGTACGC	8930 TCGTATCACG AGCATAGTGC	8940 CGAACTACGG GCITGATGCC
8950 GGCATACGCC CCGTATGCCG	8960 GTGTTGATGG CACAACCTACC	8970 CTACATCGCA GATGTAGCGT	8980 AAGAAAGTCC TTCTTTCAGG	8990 CTAGTGTAC GATCACAATG	9000 ATCGATACAG TAGCTATGTC
9010 TGCCGTGACA ACGGCACTGT	9020 GCCGTGGCCC CGGCACCGGG	9030 TGCAGCTCAT ACGTGAGTA	9040 GCCTGTTGAG CGGACAACTC	9050 ATCGTCCGCA TAGCAGGCGT	9060 AGCTAGATCA TCGATCTAGT
9070 GTCGGACTGG CAGCCTGACC	9080 GTGCGGGGTG CACGCCCCAC	9090 CCTGGATCGT GGACCTAGCA	9100 GTCAGAGACT CAGTCTCTGA	9110 TTTCCAACCTA AAAGGTTGAT	9120 GCGACCCCAA CGCTGGGGTT
9130 AGGAGTTTGG TCCTCAAACC	9140 AGCGACGATG TCGCTGCTAC	9150 ACTCCTCGAT TGAGGAGCTA	9160 GGGTGGAAGT CCCACCTTCA	9170 GATGATTGAT CTACTAACTA	9180 GATGAGAACC CTACTCTTGG

Fig. 11

9190	9200	9210	9220	9230	9240
TGACAAGAAA ACTGTTCTTT	GACGAGAGAG CTGCTCTCTC	AAATTTAGAG TTTAAATCTC	CTGTCAATTGT GACAGTAACA	AGAATTAGTC TCTTAATCAG	TAGATTCCTG ATCTAAGGAC
9250	9260	9270	9280	9290	9300
ATAATAAACA TATTATTTGT	GTATCGATTT CATAGCTAAA	TGAAACCTAA ACTTTGGATT	TTGACGTGTG AACTGCACAC	ATCGATTTTT TAGCTAAAAA	AAACCTCTGT TTTGGAGACA
9310	9320	9330	9340	9350	9360
GTTGTGTGAT CAACACACTA	TGATTGGTAT ACTAACCATA	GTGGGGGGAT CACCCCCCTA	CCGATTTCAA GGCTAAAGTT	AGGGGGGTAC TCCCCCATG	TTATCGGGAA AATAGCCCTT
9370	9380	9390	9400	9410	9420
TTGATGTGTC AACTACACAG	ATGGACGCAG TACCTGCGTC	TTTTGAGCGA AAAACCTCGCT	TTTTCCGGGA AAAAGGCCCT	ATACCGGATA TATGGCCTAT	TTACGAATTA AATGCTTAAT
9430	9440	9450	9460	9470	9480
CTGGTAGTGA GACCATCACT	CGTAGATAAT GCATCTATTA	AAAATTATAA TTTTAATATT	TGCGATTAAT ACGCTAATTA	TTTTGGTGC AAAACCACGC	TTGATTATTT AACTAATAAA
9490	9500	9510	9520	9530	9540
TTTTAGCATA AAAATCGTAT	TGTGTATCAT ACACATAGTA	TATGAGGTGA ATACTCCACT	ATGGAACAGA TACCTTGTCT	ATTACGCTGC TAATGCGACG	AGATGTCTTC TCTACAGAAG
9550	9560	9570	9580	9590	9600
ATAGAAAATG TATCTTTTAC	GCCGCCTAAT CGGCGGATTA	AAAATTATAT TTTTAATATA	TGGGTAATTA ACCCATTAAT	TTGGCTTCAT AACCGAAGTA	CGCGATCCCA GCGCTAGGGT
9610	9620	9630	9640	9650	9660
GAGGGCCCGG CTCCCCGGCC	ATGCGATAAA TACGCTATTT	AATGAACATT TTACTTGTAA	TATTGTATCC ATAACATAGG	AGACGGAAGG TCTGCCTTCC	AAACCGCCTG TTTGGCGGAC
9670	9680	9690	9700	9710	9720
GACCTGGAGT CTGGACCTCA	ATGTTTATCG TACAAATAGC	CCCGATCACC GGGCTAGTGG	TCCTCTCAAA AGAAGAGTTT	ATGGTTAGAC TACCAATCTG	AAACACAACG TTTGTGTTGC
9730	9740	9750	9760	9770	9780
ATAATAGGTG TATTATCCAC	GTATAATGTT CATATTACAA	AACATAACGA TTGTATTGCT	AATCACCAGG TTAGTGGTCC	ACCGAGACGA TGGCTCTGCT	ATAAATATAA TATTTATATT
9790	9800	9810	9820	9830	9840
CCTTGATAGG GGAACATATC	TGTTAGAGGA ACAATCTCCT	TAATATTTAA ATTATAAAAT	TGTATGTTTT ACATACAAAA	CAAACAGACA GTTTGTCTGT	AGTTCGTTAA TCAAGCAATT
9850	9860	9870	9880	9890	9900
AACAAAATAT TTGTTTTATA	TACAGTATGT ATGTCATACA	GTTTAATATG CAAATTATAC	GTGCTAACAT CACGATTGTA	GGTTGCACCA CCAACGTGGT	TCCGGTTTCA AGGCCAAAGT
9910	9920	9930	9940	9950	9960
AACTCGCATA TTGAGCGTAT	TCAATCTGTT AGTTAGACAA	ATCGGTACGA TAGCCATGCT	CACCTGTCAT GTGGACAGTA	TAATCGCATA ATTAGCGTAT	TATGTTACTT ATACAATGAA
9970	9980	9990	10000	10010	10020
ACCATATGTC TGGTATACAG	CCCTAGCCGT GGGATCGGCA	CCATGTTTTA GGTACAAAAT	GAACTAGAAG CTTGATCTTC	ATTACGACAG TAATGCTGTC	GCGCTGCCGT CGCGACGGCA
10030	10040	10050	10060	10070	10080
TGCAACAACC ACGTTGTGTG	AAATTCCTGT TTTAAGACAA	GAATACCCTG CTTATGGGAC	CCGGTCGGAA GGCCAGCCTT	CCGAATTGCT GGCTTAACGA	TAAGCCAATC ATTCCGTTAG
10090	10100	10110	10120	10130	10140
GCAGCGAGCG CGTCGCTCGC	AAAGCTGCAA TTTCGACGTT	TCGTCAGGAA AGCAGTCCTT	GTGCTGCGTA CACGACCGAT	TTTTAAAGGA AAAATTTCTT	CAAGGGAACC GTTCCCTTGG
10150	10160	10170	10180	10190	10200
AAGTGTCCTC TTCACAGAGT	ATCCTAACGC TAGGATTGCG	GCAAGCCGTG CGTTCGGCAC	CGTCGTCACA GCAGCAGTGT	TCAACCGGCT AGTTGGCCGA	ATTTTTTCGG TAAAAAGCC

Fig. 1J

11230 ACACGGCGAG TGTGCCGCTC	11240 ACCGACTTTT TGGCTGAAAA	11250 ACATGAACTG TGTACTTGAC	11260 GACGCTGCGT CTGCGACGCA	11270 CGCAGTCAGA GCGTCAGTCT	11280 CCCACCTACCT GGGTGATGGA
11290 GGAGGAGATG CCTCCTCTAC	11300 GCCTTACAGG CGGAATGTCC	11310 TGGAGATTCT ACCTCTAAGA	11320 AAAACCCCGC TTTTGGGGCG	11330 GGCGTACGTC CCGCATGCAG	11340 ACCGCGCTAT TGGCGCGATA
11350 TATCCACCAT ATAGGTGGTA	11360 CCGAAGCTAC GGCTTCGATG	11370 AGCCGGGCGT TCGGCCCCGA	11380 TGGCCTGTGG ACCGGACACC	11390 ATAGATTCT TATCTAAAGA	11400 GCGTGTACCG CGCACATGGC
11410 CTACAACGCG GATGTTGCGC	11420 CGCCTGACCC GCGGACTGGG	11430 GCGGCTACGT CGCCGATGCA	11440 ACGATACACC TGCTATGTGG	11450 CTGTACACGA GACAGTGGCT	11460 AAGCGCGCTT TTCGCGCGAA
11470 GCCCCGAAAA CGGGCGTTTT	11480 GCAGAGGGTT CGTCTCCCAA	11490 GGCTGGTGTG CCGACCACAG	11500 ACTAGACAGA TGATCTGTCT	11510 TTCATCGTGC AAGTAGCAGC	11520 AGTACCTCAA TCATGGAGTT
11530 CACATTGCTG GTGTAACGAC	11540 ATTACAATGA TAATGTTACT	11550 TGGCGGCGAT ACCGCCGCTA	11560 ATGGGCTCGC TACCCGAGCG	11570 GTTTTGATAA CAAAACTATT	11580 CCTACCTGGT GGATGGACCA
11590 GTGCGGCGGT CAGCGCCGCA	11600 CGGTAGAGGC GCCATCTCCG	11610 TTGCGGAAAC AACGCCTTTG	11620 CACGTCTCTG GTGCAGGAGC	11630 TCACACGTCG AGTGTGCAGC	11640 TTCGCGGACA AAGCGCCTGT
11650 TAGCAAGAAA ATCGTTCTTT	11660 TCCACGTCGC AGGTGCAGCG	11670 CACATCTCGA GTGTAGAGCT	11680 GAATGCCGGC CTTACGCGCC	11690 CTTGCGGGGT GAACGCCCCA	11700 CCCCTTCGCG GGGGAAGCGC
11710 CAACATTCTT GTTGTAAGGA	11720 GGCCCTGGTC CCGGGACCAG	11730 GCGTTGCGGT CGCAAGCCCA	11740 TGCTGCTTCA ACGACGAAGT	11750 GATAGACCTC CTATCTGGAG	11760 AGCGACGCTA TCGCTGCGAT
11770 CGAATGTGAC GCTTACACTG	11780 CAGCAGCACA GTCGTCTGT	11790 AAAGTCCCTA TTTCAGGGAT	11800 CTAGCACCAG GATCGTGGTC	11810 CAACAGAAAT GTTGTCTTTA	11820 AACGTGACAA TTGCAGCTGT
11830 ACGCCACGAG TGCGGTGCTC	11840 TAGCGGACCC ATCGCCTGGG	11850 ACAACCGGGA TGTTGGCCCT	11860 TCAACATGAC AGTTGTACTG	11870 CACCACCCAC GTGGTGGGTG	11880 GAGTCTTCCG CTCAGAAGGC
11890 TTCACAACGT AAGTGTGCA	11900 GCGCAATAAC CGCGTTATTG	11910 GAGATCATGA CTCTAGTACT	11920 AAGTGCTGGC TTCACGACCG	11930 TATCCTCTTC ATAGGAGAAG	11940 TACATCGTGA ATGTAGCACT
11950 CAGGCACCTC GTCCGTGGAG	11960 CATTTTCAGC GTAAAAGTCG	11970 TTCATAGCGG AAGTATCGCC	11980 TACTGATCGC ATGACTAGCG	11990 GGTAGTTTAC CCATCAAATG	12000 TCCTCGTGT AGGAGCACAA
12010 GCAAGCACCC CGTTCTGTTG	12020 GGGCCGCTTT CCCCGCGAAA	12030 CGTTTCGCCG GCAAAGCGGC	12040 ACGAAGAGGC TGCTTCTCCG	12050 CGTCAACCTG GCAGTTGGAC	12060 TTGGACGACA AACCTGCTGT
12070 CGGACGACAG GCCTGCTGTC	12080 TGGCGGCAGC ACCGCCGTCG	12090 AGCCCGTTTG TCGGGCAAAC	12100 GCAGCGGTTT CGTCGCCAAG	12110 CCGACGAGGT GGCTGCTCCA	12120 TCTCAGATCC AGAGTCTAGG
12130 CCGCCGGATT GGCGGCCTAA	12140 TTGTTCTCTG AACAAGGAGC	12150 AGCCCTTATC TCGGGAATAG	12160 AGCGGTTGGA TCGCCAACCT	12170 AACTCGGGAC TTGAGCCCTG	12180 TGGGACGAGG ACCCTGCTCC
12190 AGGAGGAGGC TCCTCCTCCG	12200 GTCCGCGGCC CAGGCGCCGG	12210 CGCGAGCGCA GCGCTCGCGT	12220 TGAAACATGA ACTTTGTACT	12230 TCCTGAGAAC AGGACTCTTG	12240 GTCATCTATT CAGTAGATAA

Fig. 1L

12250	12260	12270	12280	12290	12300
TCAGAAAGGA	TGGCAACTTG	GACACGTCGT	TCGTGAATCC	CAATTATGGG	AGAGGCTCGC
AGTCTTTTCT	ACCGTTGAAC	CTGTGCAGCA	AGCACTTAGG	GTTAATACCC	TCTCCGAGCG
12310	12320	12330	12340	12350	12360
CTTTGACCAT	CGAATCTCAC	CTCTCGGACA	ATGAGGAGGA	CCCCATCAGG	TACTACGTTT
GAAACTGGTA	GCTTAGAGTG	GAGAGCCTGT	TACTCCTCCT	GGGGTAGTCC	ATGATGCAAA
12370	12380	12390	12400	12410	12420
CGGTGTACGA	TGAACTGACC	GCCTCGGAAA	TGGAAGAACC	TTCGAACAGC	ACCAGCTGGC
GCCACATGCT	ACTTGACTGG	CGGAGCCTTT	ACCTTCTTGG	AAGCTTGTCG	TGGTCGACCG
12430	12440	12450	12460	12470	12480
AGATTCCCAA	ACTAATGAAA	GTTGCCATGC	AACCCGTCTC	GCTCAGAGAT	CCCGAGTACG
TCTAAGGGTT	TGATTACTTT	CAACGGTACG	TTGGGCAGAG	CGAGTCTCTA	GGGCTCATGC
12490	12500	12510	12520	12530	12540
ACTAGGCTTT	TTTTTTTGTC	TTTCGGTTCC	AACTCTTTCC	CCGCCCCATC	ACCTCGCCTG
TGATCCGAAA	AAAAAAACAG	AAAGCCAAGG	TTGAGAAAGG	GGCGGGGTAG	TGGAGCGGAC
12550	12560	12570	12580	12590	12600
TACTATGTGT	ATGATGTCTC	ATAATAAAGC	TTTCTTTTCTC	AGTCTGCAAC	ATGCAGCTGT
ATGATACACA	TACTACAGAG	TATTATTTTCG	AAAGAAAGAG	TCAGACGTTG	TACGTCGACA
12610	12620	12630	12640	12650	12660
GTCGGGTGTG	GCTGTCTGTT	TGTCGTGTGCG	CCGTGGTGCT	GGGTCAGTGC	CAGCGGGAAA
CAGCCCACAC	CGACAGACAA	ACAGACACGC	GGCACCACGA	CCCAGTCACG	GTGCGCCCTTT
12670	12680	12690	12700	12710	12720
CCGCGGAAAA	AAACGATTAT	TACCGAGTAC	CGCATTACTG	GGACGCGTGC	TCTCGCGCGC
GGCGCCTTTT	TTTGCTAATA	ATGGCTCATG	GCGTAATGAC	CCTGCGCACG	AGAGCGCGCG
12730	12740	12750	12760	12770	12780
TGCCCCACCA	AACCCGTTAC	AAGTATGTGG	AACAGCTCGT	GGACCTCACG	TTGAACTACC
ACGGGCTGGT	TTGGGCAATG	TTCATACACC	TTGTGAGCA	CCTGGAGTGC	AACTTGATGG
12790	12800	12810	12820	12830	12840
ACTACGATGC	GAGCCACGGC	TTGGACAAC	TTGACGTGCT	CAAGAGGTGA	GGGTACGCGC
TGATGCTACG	CTCGGTGCCG	AACCTGTTGA	AACTGCACGA	GTTCCTCCACT	CCCATGCGCG
12850	12860	12870	12880	12890	12900
TAAAGGTGCA	TGACAACGGG	AAGGTAAGGG	CGAACGGGTA	ACGGCTAAGT	AACCGCATGG
ATTTCCACGT	ACTGTTGCC	TTCCATTCCC	GCTTGCCCAT	TGCCGATTCA	TTGGCGTACC
12910	12920	12930	12940	12950	12960
GGTATGAAAT	GACGTTTGGA	ACCTGTGCTT	GCAGAAATCAA	CGTGACCCGAG	GTGTGCTTGC
CCATACTTTA	CTGCAAACCT	TGGACACGAA	CGTCTTAGTT	GCACTGGCTC	CACAGCAACG
12970	12980	12990	13000	13010	13020
TCATCAGCGA	CTTTAGACGT	CAGAACCCTC	GCGGCGGCAC	CAACAAAAGG	ACCACGTTCA
AGTAGTCGCT	GAAATCTGCA	GTCTTTGGCAG	CGCCGCCGTG	GTTGTTTTTCC	TGGTGCAAGT
13030	13040	13050	13060	13070	13080
ACGCCGCCCG	TTGCTGCGCG	CCACACGCCC	GGAGCCTCGA	GTTTCAGCGTG	CGGCTCTTTG
TGCGGCGGCC	AAGCGACCGC	GGTGTGCGGG	CCTCGGAGCT	CAAGTCGCAC	GCCGAGAAAC
13090	13100	13110	13120	13130	13140
CCAACTAGCC	TGCGTCAACG	GAAATAATAT	GCTGCGGCTT	CTGCTTCGTC	ACCACTTTCA
GGTTGATCGG	ACGCAGTGCC	CTTTATTATA	CGACGCCGAA	GACGAAGCAG	TGGTGAAAGT
13150	13160	13170	13180	13190	13200
CTGCCTGCTT	CTGTGCGCGG	TTTGGGCAAC	GCCCTGTCTG	GCGTCTCCGT	GGTCGACGCT
GACGGACGAA	GACACGCGCC	AAACCCGTTG	CGGGACAGAC	CGCAGAGGCA	CCAGCTGCGA
13210	13220	13230	13240	13250	13260
AACGGCAAAC	CAGAATCCGT	CCCCGCCATG	GTCTAAACTG	ACGTATTCCA	AACCGCATGA
TTGCCGTTTG	GTCTTAGGCA	GGGGCGGTAC	CAGATTTGAC	TGCATAAGGT	TTGGCGTACT

Fig. 1M

13270	13280	13290	13300	13310	13320
CGCGGCGACG	TTTTACTGTC	CTTTTCTCTA	TCCCTCGCCC	CCACGGTCCC	CCTTGCAATT
GCGCCGCTGC	AAAATGACAG	GAAAAGAGAT	AGGGAGCGGG	GGTGCCAGGG	GGAACGTTAA
13330	13340	13350	13360	13370	13380
CTCGGGGTTT	CAGCAGGTAT	CAACGGGTCC	CGAGTGTGCG	AACGAGACCC	TGTATCTGCT
GAGCCCCAAG	GTCGTCCATA	GTTGCCCAGG	GCTCACAGCG	TTGCTCTGGG	ACATAGACGA
13390	13400	13410	13420	13430	13440
GTACAACCGG	GAAGGCCAGA	CCTTGGTGGA	GAGAAGCTCC	ACCTGGGTGA	AAAAGGTGAT
CATGTTGGCC	CTTCCGGTCT	GGAACCACCT	CTCTTCGAGG	TGGACCCACT	TTTTCCACTA
13450	13460	13470	13480	13490	13500
CTGGTATCTG	AGCGGTGCGA	ACCAGACCAT	CCTCCAACGG	ATGCCCCAAA	CGGCTTCGAA
GACCATAGAC	TCGCCAGCGT	TGGTCTGGTA	GGAGGTTGCC	TACGGGGTTT	GCCGAAGCTT
13510	13520	13530	13540	13550	13560
ACCGAGCGAC	GGAAACGTGC	AGATCAGCGT	GGAAGACGCC	AAGATTTTTG	GAGCGCACAT
TGGCTCGCTG	CCTTTGCACG	TCTAGTCGCA	CCTTCTGCGG	TTCTAAAAAC	CTCGCGTGTA
13570	13580	13590	13600	13610	13620
GGTGCCCAAG	CAGACCAAGC	TGCTACGCTT	CGTCGTCAAC	GATGGCACGC	GTTATCAGAT
CCACGGGTTC	GTCTGGTTTC	ACGATGCGAA	GCAGCAGTTG	CTACCGTGCG	CAATAGTCTA
13630	13640	13650	13660	13670	13680
GTGTGTGATG	AAGCTGGAGA	GCTGGGCCCA	CGTCTTCCGG	GACTACAGCG	TGTCTTTTCA
CACACACTAC	TTTCGACCTCT	CGACCCGGGT	GCAGAAGGCC	CTGATGTGCG	ACAGAAAAGT
13690	13700	13710	13720	13730	13740
GGTGCGATTG	ACGTTTCACCG	AGGCCAATAA	CCAGACTTAC	ACCTTCTGTA	CCCATCCCCA
CCACGCTAAC	TGCAAGTGCG	TCCGGTTATT	GGTCTGAATG	TGGAAGACAT	GGGTAGGGTT
13750	13760	13770	13780	13790	13800
TCTCATCAAT	TGAGCCCGTC	GCGCGCGCAG	GGAATTTTGA	AAACCGCGCG	TCATGAGTCC
AGAGTAGTAA	ACTCGGGCAG	GCGCGCGCTC	CCTTAAAACT	TTTGGCGCGC	AGTACTCAGG
13810	13820	13830	13840	13850	13860
CAAAGACCTG	ACGCCGTTCT	TGACGACGTT	GTGGCTGCTA	TTGGGTCACA	GCCGCGTGCC
GTTTCTGGAC	TGCGGCAAGA	ACTGCTGCAA	CACCGACGAT	AACCCAGTGT	CGGCGCACGG
13870	13880	13890	13900	13910	13920
GCGGGTGCGC	GCAGAAGAAT	GTTGCGAATT	CATAAACGTC	AACCACCCGC	CGGAACGCTG
CGCCACGCG	CGTCTTCTTA	CAACGCTTAA	GTATTTGCAG	TTGGTGGGCG	GCCTTGCGAC
13930	13940	13950	13960	13970	13980
TTACGATTTT	AAAATGTGCA	ATCGCTTCAC	CGTCGCGTAC	GTATTTTCAT	GATTGTCTGC
AATGCTAAAG	TTTTACACGT	TAGCGAAGTG	GCAGCGCATG	CATAAAAGTA	CTAACAGACG
13990	14000	14010	14020	14030	14040
GTTCTGTGGT	GCGTCTGGAT	TTGTCTCTCG	ACGTTTCTGA	TAGCCATGTT	CCATCGACGA
CAAGACACCA	CGCAGACCTA	AACAGAGAGC	TGCAAAGACT	ATCGGTACAA	GGTAGCTGCT
14050	14060	14070	14080	14090	14100
TCCTCGGGAA	TGCCAGAGTA	GATTTTTCATG	AATCCACAGG	CTGCGGTGTC	CGGACGGCGA
AGGAGCCCTT	ACGGTCTCAT	CTAAAAGTAC	TTAGGTGTCC	GACGCCACAG	GCCTGCCGCT
14110	14120	14130	14140	14150	14160
AGTCTGCTAC	AGTCCCAGAG	AAACGGCTGA	GATTGCGGGG	ATCGTCACCA	CCATGACCCA
TCAGACGATG	TCAGGGCTCT	TTTGGCCGACT	CTAAGCGCCC	TAGCAGTGGT	GGTACTGGGT
14170	14180	14190	14200	14210	14220
TTCAATGACA	CGCCAGGTG	TACACAACAA	ACTGACGAGC	TGCAACTACA	ATCCGTAAGT
AAGTAACTGT	GCGGTCCAGC	ATGTGTTGTT	TGACTGCTCG	ACGTTGATGT	TAGGCATTCA
14230	14240	14250	14260	14270	14280
CTCTTCTCTG	AGGGCCTTAC	AGCCTATGGG	AGAGTAAGAC	AGAGAGGGAC	AAAACATCAT
GAGAAGGAGC	TCCCCGAATG	TCGGATACCC	TCTCATTCTG	TCTCTCCCTG	TTTTGTAGTA

Fig. 1N

14290	14300	14310	14320	14330	14340
TAAAAAATA	AGTCTAATTT	CACGTTTTGT	ACCCCCCTTC	CCCTCCGTGT	TGTAGCCCAT
ATTTTTTTTT	TCAGATTAAA	GTGCAAAACA	TGGGGGGAAG	GGGAGGCACA	ACATCGGGTA
14350	14360	14370	14380	14390	14400
CGGCCGCGGC	GATCTCCTAG	TAACACTCGT	CCGACACTTC	CACCATCTCC	AGCTCGGCCG
GCCGGCGCCG	CTAGAGGATC	ATTGTGAGCA	GGCTGTGAAG	GTGGTAGAGG	TCGAGCCGGC
14410	14420	14430	14440	14450	14460
GCGGTTTCGGC	ATCCTCTACC	AGCGGCGTCG	TCTCATCTTT	GCCGCAGCAG	CGGACGCACA
CGCCAAGCCG	TAGGAGATGG	TCGCCGCAGC	AGAGTAGAAA	CGGCGTCGTC	GCCTGCGTGT
14470	14480	14490	14500	14510	14520
CCTTCTCCAG	GCAGAACGCC	ACCAGCTGCC	GCCGAACGTA	CCACAGGTAC	ACGTGCAGAC
GGAAGAGGTC	CGTCTTGCGG	TGGTCGACGG	CGGCTTGTCAT	GGTGTCCATG	TGCACGTCTG
14530	14540	14550	14560	14570	14580
CTGCGAACAG	GACTACGGAG	GTCATGACCA	CCACGACGCA	CACGGGAATC	CAGGGATCGA
GACGCTTGTC	CTGATGCCTC	CAGTACTGGT	GGTGCTGCGT	GTGCCCTTAG	GTCCCTAGCT
14590	14600	14610	14620	14630	14640
GATTGTTGCT	GGAATCATAG	GCTATCGCCA	CCGACGTGCC	CGCGTCTGTC	TCACCGCCGC
CTAACAACGA	CCTTGAGTAC	CGATAGCGGT	GGCTGCACGG	GCGCAGACAG	AGTGGCGGCG
14650	14660	14670	14680	14690	14700
TCGCCCGATG	TCGCGCGGCT	TGTTATACGC	TAGCCCGTCG	CCGCCTCGGG	GCACGGTGCC
AGCGGGCTAC	AGCGCGCCGA	ACAATATGCG	ATCGGGCAGC	GGCGGAGCCC	CGTGCCACGG
14710	14720	14730	14740	14750	14760
CTCCTACCCA	CGTAACCTCC	TCCGTGACTT	AAAGTCGCGT	GTGGTAGATC	TCCTGCTCCG
GAGGATGGGT	GCATTGAAGG	AGGCACTGAA	TTTCAGCGCA	CACCATCTAG	AGGACGAGGC
14770	14780	14790	14800	14810	14820
TGGACGAACC	GTCCGGCAGG	ATAGCGGTTA	AGGATTCGGT	GCTAAGGCCG	TGTCGCCAAC
ACCTGCTTGG	CAGGCCGTCC	TATCGCCAAT	TCCTAAGCCA	CGATTCCGGC	ACAGCGGTTG
14830	14840	14850	14860	14870	14880
GTCGAATGCT	ACGTTGCAAC	AGCTTCGACG	GACGGCCATC	CCCTCTCTCA	TCGCAATAAT
CAGCTTACGA	TGCAACGTTG	TCGAAGCTGC	CTGCCGGTAG	GGGAGAGAGT	AGCGTTATTA
14890	14900	14910	14920	14930	14940
AAAACACCAG	CAGCGCGCAC	GACGCGATCA	CGGTGACACC	CATGATTAGA	CCCACGCAGA
TTTTGTGGTC	GTCGCGCGTG	CTGCGCTAGT	GCCACTGTGG	GTACTAATCT	GGGTGCGTCT
14950	14960	14970	14980	14990	15000
TAGCCAGCCC	CGCTAGCGTA	TCTAGCGCCA	TCCCGTTTCG	TCCCGTTGTC	TCCTGAGCGA
ATCGGTCGGG	GCGATCGCAT	AGATCGCGGT	AGGGCAAGCG	AGGGCAACAG	AGGACTCGCT
15010	15020	15030	15040	15050	15060
AGCAACTTCT	CGGTCCCCGT	TTTCAACAGT	TTTTGTTTTCC	TTCTCCGCGA	CTAGATGTTA
TCGTTGAAGA	GCCAGGGGCA	AAAGTTGTCA	AAAACAAAGG	AAGAGGCGCT	GATCTACAAT
15070	15080	15090	15100	15110	15120
ACGCGCGCGG	TCTTTCCGGC	CGTGCTCTAC	CTCCTGGCGC	TTGTCTGCTG	GGTTGAGATG
TGCGGGCGCC	AGAAAGGCCG	GCACGAGATG	GAGGACCGCG	AACAGCAGAC	CCAACTCTAC
15130	15140	15150	15160	15170	15180
TTCTGCCTCG	TCGCCGTAGC	CGTCGTCGAG	CGCGAGATCG	CCTGGGCGCT	GCTGCTGCGG
AAGACGGAGC	AGCGGCATCG	GCACGAGCTC	GCGCTCTAGC	GGACCCGCGA	CGACGACGCC
15190	15200	15210	15220	15230	15240
ATGCTGGTCG	TTGGCCTGAT	GGTGAAGTTC	GGCGCCGCGG	CCGCTTGGAC	CTTCGTGCGT
TACGACCAGC	AACCGGACTA	CCACCTTCAG	CCGCGGCGGC	GGCGAACCTG	GAAGCACGCA
15250	15260	15270	15280	15290	15300
TGTCTTGCCCT	ATCAGCGCTC	CTTCCCCGTG	CTTACGGCCT	TCCCCTGAAA	CCCACGTTAA
ACAGAACGGA	TAGTCGCGAG	GAAGGGGCAC	GAATGCCGGA	AGGGGACTTT	GGGTGCAATT

Fig. 10

15310	15320	15330	15340	15350	15360
CCGACCGTCC	CAAAAACGCC	GGTGTTAACA	CAGGAAAAAA	AGAAACCACG	CAGGAACCGC
GGCTGGCAGG	GTMTTTCGCG	CCACAATTGT	GTCTTTTMTT	TCCTTTGGTG	GTCTTTGGCG
15370	15380	15390	15400	15410	15420
GCAGGAACCA	CGCGGAACAT	GGGACACTAT	CTGGAAATCC	TGTTCAACGT	CATCGTCTTC
CGTCCCTGGT	GCGCCTTGTA	CCCTGTGATA	GACCTTTAGG	ACAAGTTGCA	GTAGCAGAAG
15430	15440	15450	15460	15470	15480
ACTCTGCTGC	TCGGCGTCAT	GGTCAGTATC	GTCGCTTGGT	ACTTCACGTG	AACCACCGTC
TGAGACGACG	AGCCGCAGTA	CCAGTCATAG	CAGCGAACCA	TGAAGTGCAC	TTGGTGGCAG
15490	15500	15510	15520	15530	15540
GTCCCGGTTT	AAAAACCATC	ATCGACGGCC	GTTATAAAGC	CACCCGGACA	CGCGCCGCGG
CAGGGCCAAA	TTTTTGGTAG	TAGCTGCCGG	CAATATTTTC	GTGGGCCTGT	GCGCGGCGCC
15550	15560	15570	15580	15590	15600
CACTTGCCCTA	CGGCGCTGCT	TCAGGGAAC	TCCTCTTCCT	TCTGCTCTTC	CTCCTTCACC
GTGAACGGAT	GCCGCGACGA	AGTCCCTTTG	AGGAGAAGGA	AGACGAGAAG	GAGGAAGTGG
15610	15620	15630	15640	15650	15660
GCAGGGATCG	TTTCCCTCGA	CCAGGGACTC	GCCGAAGCAA	CCGCCGGAGC	AACCTGGAGG
CGTCCCTAGC	AAAGGGAGCT	GGTCCCTGAG	CGGCTTCGTT	GGCGGCCTCG	TTGGACCTCC
15670	15680	15690	15700	15710	15720
AGTCGCGGCA	TGACGGCGCC	CAAGTGTGTC	ACCACCAGTA	CTTATCTGGT	CAAGACCAAG
TCAGCGCCGT	ACTGCCGCGG	GTTACACAG	TGGTGGTCAT	GAATAGACCA	GTTCTGGTTC
15730	15740	15750	15760	15770	15780
GAACAGCCCT	GGTGGCCCGA	CAACGCCATC	AGGAGATGGT	GGATCAGTGT	TGCTATCGTC
CTTGTTCGGA	CCACCGGGCT	GTTGCGGTAG	TCCTCTACCA	CCTAGTCACA	ACGATAGCAG
15790	15800	15810	15820	15830	15840
ATCTTCATCG	GAGTCTGTCT	GGTGGCCCTG	ATGTACTTTA	CGCAGCAGCA	GGCACGCAGC
TAGAAGTAGC	CTCAGACAGA	CCACCGGGAC	TACATGAAAT	GCGTCGTCGT	CCGTGCGTCG
15850	15860	15870	15880	15890	15900
GGGAGCAGCA	GCGGCTAGAC	AAGTCTCTGG	CGGCTACAGC	TCCAAGCGCC	GTAGCCGGGC
CCCTCGTCGT	CGCCGATCTG	TTGAGAGACC	GCCGATGTCT	AGGTTTCGCG	CATCGGCCCC
15910	15920	15930	15940	15950	15960
CGCCTGCCGA	TCGCGACGTC	GTGGACCATC	GAACAGAGAC	TCACGCGTAC	GAGACCCCGA
GCGGACGGCT	AGCGCTGCAG	CACCTGGTAG	CTTGTCTCTG	AGTGCGCATG	CTCTGGGGCT
15970	15980	15990	16000	16010	16020
GGTACGCCAC	GCGGTGCCTA	ACGCGGTATA	CCACACCCGT	ACGGTCTGCA	GTGCGGCGTA
CCATGCGGTG	CGCCACGGAT	TGCGCCATAT	GGTGTGGGCA	TGCCAGACGT	CACGCCGCAT
16030	16040	16050	16060	16070	16080
CAACGTGTGG	AAAACGCGTT	GCGTCGCAGA	GTCCGCCACG	TTCTGTCTTT	GTGCTCTCCC
GTTCACACAC	TTTTTGCGAA	CGCAGCGTCT	CAGGCGGTGC	AAGGACAGAA	CAGCGAGGGG
16090	16100	16110	16120	16130	16140
AATCGTCTCC	CGCACACCCC	CCGCGACACC	CAGAGGGCGG	GTGAGCCAAG	TATTCTTAAG
TTAGCAGAGG	GCGTGTGGGG	GGCGCTGTGG	GTCTCCCGCC	CACTCGGTTC	ATAAGAATTC
16150	16160	16170	16180	16190	16200
GCCGTCTCTT	GTTCCATAGC	CCATAAATTG	TTGATTCCGG	AGCTCGTTGG	CGCGGAAATA
CGGCAAGAAA	CAAGGTATCG	GGTATTTAAC	AACTAAGGCC	TCGAGCAACC	GCGCCTTTAT
16210	16220	16230	16240	16250	16260
GCCGGATAAG	GGGAGCAACA	ACCGTTGGCG	AAAGCCGTCC	CGCTCATTCA	GTCCGGGTTT
CAGCCTATTG	CCCTCGTTGT	TGGCAACCGC	TTTCGGCAGG	GCGAGTAAGT	CAGGCCCAAA
16270	16280	16290	16300	16310	16320
CGCGTCCAGT	CGGACGTGTG	ACCGTTGGGC	AACGGAACGG	CGTTTCACTG	CCAAAATCGT
GCGCAGGTCA	GCCTGCACAC	TGGCAACCCG	TTGCCTTGCC	GCAAAGTGAC	GGTTTTAGCA

Fig. 1P

16330	16340	16350	16360	16370	16380
ATCGGGTAGT	GTACGAGACG	TCGGCGGTGC	AGAATGCGAC	TCGCGGCGTA	GCTCGCCGTC
TAGCCCATCA	CATGCTCTGC	AGCCGCCACG	TCTTACGCTG	AGCGCCGCAT	CGAGCGGCAG
16390	16400	16410	16420	16430	16440
GCTATGCGGC	TCGTCGCCGT	GTGGCGCGGC	CTGGCCGGCT	GTCTGCGTCC	AGATCTGTTG
CGATACGCCG	AGCAGCGGCA	CACCGCGCCG	GACCGGCCGA	CAGACGCAGG	TCTAGACAAC
16450	16460	16470	16480	16490	16500
GCCTTTTGGT	TCCTCTGGCT	GCTGCTCGGT	GTGTGCTTTG	GTAGACGCGG	TGGCAGTTTG
CGGAAAACCA	AGGAGACCGA	CGACGACGCA	CACACGAAAC	CATCTGCGCC	ACCGTCAAAC
16510	16520	16530	16540	16550	16560
CGTCTGCGG	TAAGTGAGGA	TGTCGCCGAG	CAAACGCAC	TGCGGCGCGT	GGCGGGCACG
GCCAGACGCC	ATTCACTCCT	ACAGCGGCTC	GTTTGCGTGA	ACGCCGCGCA	CCCGCCGTGC
16570	16580	16590	16600	16610	16620
CGTGTCATTG	TAGGTTTCGTT	GCCAGATGGC	AAGTGCTGTC	AACAGCAGGC	GTGTGGGCG
GCACAGTAAC	ATCCAAGCAA	CGGTCTACCG	TTCACGACAG	TTGTGCTCCG	CAACACCCGC
16630	16640	16650	16660	16670	16680
GTCGGTGTAT	TTTTGTGGGT	TGCGGTGAGA	GTCGGCACTC	GGTGTTTTGT	GAGTCATCTC
CAGCCACATA	AAAAACACCA	ACGCCACTCT	CAGCCGTGAG	CCACAAAACA	CTCAGTAGAG
16690	16700	16710	16720	16730	16740
AACTATCTGT	GTGTCTTTTGA	GCAGCGTCCA	GAACAGCGAC	GCGACTTTGG	GGATGGCCTC
TTGATAGACA	CAACGAAACT	CGTCGCAGGT	CTTGTCGCTG	CGCTGAAACC	CCTACCGGAG
16750	16760	16770	16780	16790	16800
GTGCTCACCT	CCGCGGAGAG	CGCCGCCGGA	CCTGCTCGTC	AGCAGCGAGC	TACGCAGACG
CACGAGTGGA	GGCGCCTCTC	GCGCGGCCT	GGACGAGCAG	TCGTGCTCG	ATGCGTCTGC
16810	16820	16830	16840	16850	16860
GAATATCTGG	AGGAGAGTTA	CGTGTTCTAC	AGGAGAGCGC	GGGTCTCCGG	CGGTAACGAC
CTTATAGACC	TCCTCTCAAT	GCACACAGTG	TCCTCTCGCG	CCCAGAGGCC	GCCATTGCTG
16870	16880	16890	16900	16910	16920
GGCGGTGTG	TCGACACGTG	TGCGGCCTGT	TGTGCTCTGC	GGAAAAGTGC	CGGTCTCGGA
CCGCCACAGC	AGCTGTGCAC	ACGCCGGACA	ACACGAGACG	CCTTTTTCACG	GCCAGAGCCT
16930	16940	16950	16960	16970	16980
GACCGTGGAC	GAAAAAGAGA	ACGCAGCAGC	TACCGCTGGC	GGCGGCGGCG	TTAATGCAGC
CTGGCACCTG	CTTTTTCTCT	TGCGTCGTG	ATGGCGACCG	CCGCCGCCGC	AATTACGTG
16990	17000	17010	17020	17030	17040
CGTTGATGTT	CGACGTTGTG	AGCACTCGGA	AACAGCGGTG	AGGCAGAAGG	TCGATTCTCC
GCAACTACAA	GCTGCAACAC	TCGTGAGCCT	TTGTGCGCC	TCCGTCTTCC	AGCTAAGAGG
17050	17060	17070	17080	17090	17100
AGGGAACGAC	AGTCGATGCG	TGGTAGCCGC	AGCAGGTGAG	GTGGGGGCGG	ACAACGTGTT
TCCCTTGCTG	TCAGCTACGC	ACCATCGGCG	TCGTCCACTC	CAACCCCGCC	TGTTGCACAA
17110	17120	17130	17140	17150	17160
GCGGATGTG	GCGAGAACGT	CGTCCTCCCC	TTCTTACACG	CCCCACCCAC	CCTCGGTTGG
CGCTTAACAC	CGCTCTTGCA	GCAGGAGGGG	AAGAAGTGCC	GGGGTGGGTG	GGAGCCAACC
17170	17180	17190	17200	17210	17220
TGTTTCTTTT	TTCTTGTGTC	CTGCAGATAG	TTCCACGGAC	AGCGACGGCA	AGTCCATAAT
ACAAAGAAAA	AAGAACACAG	GACGTCTATC	AAGGTGCCTG	TCGCTGCCGT	TCAGGTATTA
17230	17240	17250	17260	17270	17280
CAGCGGTGTG	CAAGTGGTGG	AACACGACGA	AGATATCATC	GCGCCGCAGA	GTTTGTGGTG
GTCGCCACAC	GTTCAACCAC	TTGTGCTGCT	TCTATAGTAG	CGCGGCGTCT	CAAACACCAC
17290	17300	17310	17320	17330	17340
CACGGCGTTC	AAGGAAGCCC	TCTGGGATGT	GGCTCTGTTG	GAAGTGCCCG	GTTGGGCGTG
GTGCCGCAAG	TTCTTTCGGG	AGACCTTACA	CCGAGACAAC	CTTACAGCGG	CAACCCGCAC

Fig. 1Q

17350	17360	17370	17380	17390	17400
GCAGGGCTGG	AAGAGGTGGC	GCAACAGCGA	GGCCGGGCGT	CGATGGAGTG	CTGGGTCTGC
CGTCCCGACC	TTCTCCACCG	CGTTGTCTGT	CCGGCCCGCA	GCTACCTCAC	GACCCAGACG
17410	17420	17430	17440	17450	17460
GTCGGCTTCC	AGCTTGTCTG	ACTTGGCGGG	CGAGGCCGTT	GGAGAATTGG	TGGGATCGGT
CAGCCGAAGG	TGGAACAGAC	TGAACCGCCC	GCTCCGGCAA	CCTCTTAACC	ACCCTAGCCA
17470	17480	17490	17500	17510	17520
CGTCGCGTAC	GTGATCCTTG	AACGTCTGTG	GTTGGCAGCC	AGAGGTTGGG	TGTGCGAAAC
GCAGCGCATG	CACTAGGAAC	TTGCAGACAC	CAACCGTCGG	TCTCCAACCC	ACACGCTTTG
17530	17540	17550	17560	17570	17580
AGGTGTGGAA	GCCGAGGAGG	CCATGTCGCG	GCGGCGACAG	CGCATGCTGT	GGCGTATTGT
TCCACACCTT	CGGCTCCTCC	GGTACAGCGC	CGCCGCTGTC	GCGTACGACA	CCGCATAACA
17590	17600	17610	17620	17630	17640
TCTCTCGTGG	AGGCGACGGC	GAATGCAGCA	GACGGTGTTC	GATGGAGATG	GCGTGCGGGG
AGAGAGCACC	TCCGCTGCCG	CTTACGTCGT	CTGCCACAAG	CTACCTCTAC	CGCACGCCCC
17650	17660	17670	17680	17690	17700
AAGAAAGCGC	CGTGTGTGTA	GCAGACGACG	TAGGATGCGG	GACGTCGGAG	CACATGGGCC
TTCTTTTCGG	GCACAACACT	CGTCTGCTGC	ATCCTACGCC	CTGCAGCCTC	GTGTACCCGG
17710	17720	17730	17740	17750	17760
ATGTGTGGTG	GCAGATGGCG	GTGTCCGCTG	GTGTCTGCTG	CGGCAGTGCA	TAGACGAAGC
TACACACCAC	CGTCTACCGC	CACAGGCGAC	CACAGACGAC	GCCGTCACGT	ATCTGCTTCG
17770	17780	17790	17800	17810	17820
AACATGTGCG	TGTGAAGAGA	TAGAGTGTGA	GCATAGCTGC	ATGCAGCGTT	GCGTGTATAA
TTGTACAGCG	ACACTTCTCT	ATCTCACACT	CGTATCGACG	TACGTGCGAA	CGCACATATT
17830	17840	17850	17860	17870	17880
GCGGGGGGGA	TTAAGACGTT	AATAAAGAAT	AGCGGCGGTT	CTGATAGGGC	GACCGCTGAA
CGCCCCCCTT	AATTCTGCAA	TTATTTCTTA	TGCGCCGCAA	GACTATCCCG	CTGGCGACTT
17890	17900	17910	17920	17930	17940
GTGAGCTGCG	TGTGCGTGTG	GTTTGTGGAG	TCCCCGCCGC	CCCCGGTCCC	GTGTCCGCCG
CACTCGACGC	ACACGCACAC	CAAACACCTC	AGGGGCGGCG	GGGGCCAGGG	CACAGGCGGC
17950	17960	17970	17980	17990	18000
GCAAAGCCCC	CCGNTTCCGC	ACACTCCTGG	CCGCGCAACC	CTCGTCTGCT	CAAAAGCCCC
CGTTTCGGGG	GGCCNAGGCG	TGTGAGGACC	GGCGCGTTGG	GAGCAGCGAC	GTTTTTCGGG
18010	18020	18030	18040	18050	18060
CCGTCCCCGC	ACACCCCCGC	GACCGCCGGT	CCCGCGAGTC	CCCGTCCCCG	CCGCAAAAGG
GGCAGGGGCG	TGTGGGGGCG	CTGGCGGCCA	GGGCGCTCAG	GGGCAGGGGC	GGCGTTTTCC
18070	18080	18090	18100	18110	18120
CCCCCGTCCT	CGCCGCAAAC	ACCCCGCTCA	CCCCCGTCCC	TCAGNCCGGG	TCCGCGAGTC
GGGGGCAGGA	GCGGCGTTTG	TGGGGGCAGT	GGGGGCAGGG	AGTCNGGCCC	AGGCGCTCAG
18130	18140	18150	18160	18170	18180
CCCGTTCCCA	GCGTAATCCC	CGTACCCGCA	ACGNCCCGGN	CCCACCGTCG	TCCCGCACAC
GGGCAAGGGT	CGCATTAGGG	GCATGGGCGT	TGCNNGGGCN	GGGTGGCAGC	AGGGCGTGTG
18190	18200	18210	18220	18230	18240
CCCCCGTCCC	CCAGCCCGGT	GCCCAGCGTG	CGAAAAAAGC	TCCGTCCCTC	ACACCCGCGC
GGGGGCAGGG	GGTCGGGCCA	CGGGTCGCAC	GCTTTTTTTC	AGGCAGGGAG	TGTGGGCGTC
18250	18260	18270	18280	18290	18300
AAAGATCCCT	CAGCGCGGTG	AAACCCCGTC	CCCAGCGCCG	TGCCGCTGAC	AAAGACCATG
TTTCTAGGGA	GTCGCGCCAC	TTTGGGGCAG	GGGTGCGGGC	ACGGCGACTG	TTTCTGGTAC
18310	18320	18330	18340	18350	18360
GGACGACACG	CACAGGCA..
CCTGCTGTGC	GTGTCCGT..

Fig. 1R

DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I declare that:

My residence, post office address and citizenship are as stated below next to my name; I believe I am an original, first and joint inventor (if plural inventors are named below) of the subject matter which is claimed and for which a patent is sought on the invention entitled: ATTENUATION OF CYTOMEGALOVIRUS VIRULENCE the specification of which is attached hereto. I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, Section 1.56. I claim foreign priority benefits under Title 35, United States Code, Section 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed.

Prior Foreign Application(s)

Country	Application No.	Date of Filing	Priority Claimed Under 35 USC 119
			Yes <input type="checkbox"/> No <input type="checkbox"/>
			Yes <input type="checkbox"/> No <input type="checkbox"/>

I hereby claim the benefit under Title 35, United States Code 119(e) of any United States provisional application(s) listed below:

Application No.	Filing Date

I claim the benefit under Title 35, United States Code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, Section 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, Section 1.56 which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Application No.	Date of Filing	Status
		<input type="checkbox"/> Patented <input type="checkbox"/> Pending <input type="checkbox"/> Abandoned
		<input type="checkbox"/> Patented <input type="checkbox"/> Pending <input type="checkbox"/> Abandoned

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith.

Tracy J. Dunn Reg. No. 34,587

Send Correspondence to: Tracy J. Dunn AVIRON 297 Bernardo Avenue Mt. View, CA 94043	Direct Telephone Calls to: Tracy J. Dunn Reg. No. 34,587 650.919.6637 (voice) 650.919.6610 (fax)
---	---

Full Name of Inventor 1	Last Name KEMBLE	First Name George	Middle Name or Initial W.	
Residence & Citizenship	City Fremont	State/Foreign Country CALIFORNIA	Country of Citizenship USA	
Post Office Address	Post Office Address 44190 Boitano Avenue	City Fremont	State/Country CALIFORNIA	Zip Code 94539
Full Name of Inventor 2	Last Name DUKE	First Name Gregory	Middle Name or Initial M.	
Residence & Citizenship	City Palo Alto	State/Foreign Country CALIFORNIA	Country of Citizenship USA	
Post Office Address	Post Office Address 659 Forest Avenue	City Palo Alto	State/Country CALIFORNIA	Zip Code 94301
Full Name of Inventor 3	Last Name SPAETE	First Name Richard	Middle Name or Initial R.	
Residence & Citizenship	City Redwood City	State/Foreign Country California	Country of Citizenship USA	
Post Office Address	Post Office Address 720 Lakemead Way.	City Redwood City	State/Country CALIFORNIA	Zip Code 94062

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Signature of Inventor 1	Signature of Inventor 2	Signature of Inventor 3
GEORGE W. KEMBLE	GREGORY M. DUKE	RICHARD R. SPAETE
Date	Date	Date